Data Assimilation as a Problem in Optimal Tracking: Application of Pontryagin's Minimum Principle to Atmospheric Science

S. LAKSHMIVARAHAN

School of Computer Science, University of Oklahoma, Norman, Oklahoma

J. M. LEWIS

National Severe Storms Laboratory, Norman, Oklahoma, and Desert Research Institute, Reno, Nevada

D. PHAN

School of Computer Science, University of Oklahoma, Norman, Oklahoma

(Manuscript received 31 July 2012, in final form 13 October 2012)

ABSTRACT

A data assimilation strategy based on feedback control has been developed for the geophysical sciences—a strategy that uses model output to control the behavior of the dynamical system. Whereas optimal tracking through feedback control had its early history in application to vehicle trajectories in space science, the methodology has been adapted to geophysical dynamics by forcing the trajectory of a deterministic model to follow observations in accord with observation accuracy. Fundamentally, this offline (where it is assumed that the observations in a given assimilation window are all given) approach is based on Pontryagin's minimum principle (PMP) where a least squares fit of idealized path to dynamic law follows from Hamiltonian mechanics. This utilitarian process optimally determines a forcing function that depends on the state (the feedback component) and the observations. It follows that this optimal forcing accounts for the model error. From this model error, a correction to the one-step transition matrix is constructed. The above theory and technique is illustrated using the linear Burgers' equation that transfers energy from the large scale to the small scale.

1. Introduction

Data assimilation as a means of constructing the initial conditions for dynamical prediction models in meteorology has 50+ yr of history. It began in the late 1940s–early 1950s as a response to anticipation of numerical weather prediction (NWP) that began in a research mode at Princeton's Institute for Advanced Study (IAS) in 1946 [reviewed in Lynch (2006)]. By the mid-1950s, operational NWP commenced in Sweden and shortly thereafter in the United States (Wiin-Nielsen 1991). The first operational numerical weather map analysis or objective analysis as it was then called came from the work of Bergthórsson and Döös (1955)—the B–D scheme.

The pragmatic and utilitarian B–D scheme established the following guidelines that became central to development of meteorological data assimilation: 1) use of a background field that, in their case, was a combination of a forecast from an earlier time (12 h earlier) and climatology; and 2) interpolation of an ''increment'' field, the difference between the forecast and observation at the site of the observation, to grid points as a means of adjusting the background. Two optimal approaches to data assimilation came in the wake of the B–D scheme. The first was a stochastic method designed by Eliassen (1954) with further development and operational implementation by Gandin (1965) at the National Meteorological Center (NMC), United States [reviewed in Bergman (1979)]. The second was a deterministic scheme developed by Sasaki (1958, 1970a,b,c) with operational implementation by Lewis (1972) at the U.S. Navy's Fleet Numerical Weather Center (FNWC). The subsequent advancement of these two approaches became known

Corresponding author address: S. Lakshmivarahan, School of Computer Science, University of Oklahoma, 110 W Boyd St., Room DEH 230, Norman, OK 73019.
E-mail: varahan@ou.edu

DOI: 10.1175/JAS-D-12-0217.1

as three-dimensional variational data assimilation (3DVAR) and four-dimensional variational data assimilation (4DVAR), respectively. A comprehensive review of the steps that led to these developments is found in the historical paper by Lewis and Lakshmivarahan (2008). As currently practiced, both 3DVAR and 4DVAR make use of a background, a forecast from an earlier time, and thereby embrace a Bayesian philosophy (Kalnay 2003; Lewis et al. 2006).

The subject of automatic control and feedback control in particular came into prominence in the immediate post–World War II (WWII) period (Wiener 1948) when digital computers became available and control of ballistic objects such as missiles and rockets took center stage in the Cold War era (Bennett 1996; Bryson 1996). Development of mathematical algorithms to optimally track rockets and artificial satellites and to efficiently and economically change their course became a fundamental theme in control theory. One of the algorithms developed during this period became known as Pontryagin's minimum principle (PMP) (Pontryagin et al. 1962; Boltyanskii 1971, 1978; Bryson 1996, 1999). This principle, developed by Lev Pontryagin and his collaborators, is expressed in the following form: In the presence of dynamic constraints (typically differential equations of motion), find the best possible control for taking a dynamic system from one state to another. Essentially, this principle embodies the search for minimization of a cost function that contains the Euler–Lagrange search for the minimum. As will be shown in section 3, 4DVAR is a special case of PMP. We will test this methodology and concept in meteorological problems where the task will be to force the system toward observations in much the same spirit as the nudging method (Anthes 1974)—but importantly, in this case, the process is optimal (Lakshmivarahan and Lewis 2013).

In this paper we succinctly review the basis for the PMP as it applies to the determination of the optimal control/forcing by minimizing a performance functional that is a sum to two quadratic forms representing two types of energy where the given model is used as a strong constraint. The first term of this performance functional is the total energy of the error, the difference between the observations (representing truth), and model trajectory starting from an arbitrary initial condition. Minimization of this energy term has been the basis for the variational methods (Lewis et al. 2006). The second quadratic form represents the total energy in the control signal. It must be emphasized that the use of least energy to accomplish a goal is central to engineering design and distinguishes this approach from the traditional variational approaches to dynamic data assimilation. A family of optimal controls can be achieved by giving different weights to these two energy terms.

By introducing an appropriate Hamiltonian function, this approach based on PMP reduces the well-known second-order Euler–Lagrange equation to a system of two first-order canonical Hamiltonian equations, the like of which have guided countless developments in physics (Goldstein 1950, 1980). While Kuhn and Tucker (1951) extended the Lagrangian technique for equality constraints to include inequality constraints by developing the theory of nonlinear programming for static problems, Pontryagin et al. (1962) used this Hamiltonian formulation to extend the classical Euler–Lagrange formulation in the calculus of variations. This extension has been the basis for significant development of optimal control theory in the dynamical setting. The resulting theory is so general that it can handle both equality and inequality constraints on both the state and the control. Further, there is a close relationship between the PMP and Kuhn–Tucker condition. See Canon et al. (1970) for details.

Recall that the optimal control computed using the PMP forces the model trajectory toward the observations. Hence, it is natural to interpret this optimal control as the additive optimal model error correction. In an effort to further understand the impact of knowing this optimal sequence of model errors, we take PMP one step further. Given an erroneous linear model with $\mathbf{M}$ as its one-step state transition matrix, we have developed a flexible framework that consolidates the information in the optimal model error sequence into a correction matrix $\mathbf{S}$ such that the corrected model governed by $(\mathbf{M} + \mathbf{S})$ will match the optimal trajectory.

While the PMP approach to dynamic data assimilation in meteorology is new, there are conceptual and methodological similarities between this approach and the vast literature devoted to analysis of model errors. We explore some of the similarities. The contributions in the area of model error correction are broadly classified along two lines—deterministic or stochastic model and the model constraint that is strong or weak.

In a stimulating paper, Derber (1989) first postulates that the deterministic model error can be expressed as the product of an unknown temporally fixed spatial function ϕ and a prespecified time-varying function. Using the model as a strong constraint, he then extends the 4DVAR method to estimate ϕ along with the initial conditions which to our knowledge represents the first attempt to quantify the model errors using the variational framework. Griffith and Nichols (2001) again postulate that the model error evolves according to an auxiliary model with unknown parameters. By augmenting this empirical secondary model with the given

model, they then estimate both the initial condition and the parameters using the 4DVAR, using the model as a strong constraint. The PMP-based approach advocated in this paper does not rely on empirical auxiliary models.

It is also appropriate to briefly mention the earlier efforts in control theory and meteorology to account for model error. See Rouch et al. (1965), Friedland (1969), Bennett (1992), Bennett and Thorburn (1992), and Dee and da Silva (1998). In the spirit of these contributions, the work by Menard and Daley (1996) made the first attempt to relate Kalman smoothing to PMP. The primary difference between our approach and the Menard and Daley (1996) approach is that we consider a deterministic strong constraint model with time-varying errors while they develop a weak constraint stochastic model formulation with stochastic error terms with known covariance structure. Zupanski's (1997) discussion of advantages with the weak constraint formulation of the 4DVAR to assess systematic and random model errors is a meaningful complement to Menard and Daley (1996).

In section 2 we provide a robust derivation of the PMP for the general case of (autonomous) nonlinear model where observations are (autonomous) nonlinear functions of the state. The computation of the optimal control sequence in this general case reduces to solving a nonlinear two-point boundary-value problem (TPBVP). We then specialize these results for the case when both the model and observations are linear in section 3. In this case of linear dynamics, the well-known sweep method (Kalman 1963) is used to reduce the TPBVP to solve two initial-value problems. To illustrate the power of the PMP we have chosen the linear Burgers equation where the advection velocity is a sinusoidal function of the space variable—this linear model has many of the characteristics of Platzman's (1964) classic study of Burgers's nonlinear advection. Many of the key properties of this linear Burgers equation and its n-mode spectral counterpart [also known as the low-order model LOM(n)] obtained by using the standard Galerkin projection method (Shen et al. 2011) are described in section 4. Numerical experiments relating to the optimal control of LOM(4) are given in section 5. In a series of interesting papers, Majda and Timofeyev (2000, 2002) and Abramov et al. (2003) analyze the statistical properties of the solution of the n-mode spectral approximation to the nonlinear Burgers equation. Section 6 illustrates the computation of the consolidated correction matrix using the computed time series of optimal controls and the associated optimal trajectory. It is demonstrated that the uncontrolled solution of the corrected model ($\mathbf{M} + \mathbf{S}$) indeed matches the optimal trajectory of the model. Section 7 contains concluding remarks. The three appendices provide supplementary results used in derivation found in the main body of the text.

2. Minimum principle in discrete form

a. Stepwise solution of the variational problem

In this section we provide a summary of the celebrated Pontryagin minimum principle, which is based on expressing the classical Lagrangian function in terms of the Hamiltonian function. In the following we follow the developments in Athans and Falb (1966), Lewis (1986), and Naidu (2003). Let

$$\overline{\mathbf{x}}_{k+1} = \overline{\mathbf{M}}(\overline{\mathbf{x}}_k, \boldsymbol{\eta}_k) \tag{2.1}$$

be the given discrete time nonlinear model dynamics where $\overline{\mathbf{M}}$: $\mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R}^n$, $\overline{\mathbf{x}}_k \in \mathbf{R}^n$ is the state of the time invariant dynamics, and $\boldsymbol{\eta}_k \in \mathbf{R}^n$ is the given intrinsic physical forcing that is part of the model.

Pontryagin's method calls for adding an external forcing term to the given model dynamics in (2.1). Let the resulting forced dynamics be given by

$$\mathbf{x}_{k+1} = \mathbf{M}(\mathbf{x}_k, \boldsymbol{\eta}_k, \mathbf{u}_k) = \overline{\mathbf{M}}(\mathbf{x}_k, \boldsymbol{\eta}_k) + \mathbf{B}\mathbf{u}_k, \tag{2.2}$$

where $1 \le p \le n$, $\mathbf{B} \in \mathbf{R}^{n \times p}$, and $\mathbf{u}_k \in \mathbf{R}^p$ is the new control or decision vector. As an example, when $p = 1$ and $\mathbf{B} = (1, 1, \ldots, 1)^{\mathrm{T}} \in \mathbf{R}^n$, then the same (scalar) control u_k is applied to each and every component of the state vector. At the other extreme, when $p = n$ and $\mathbf{B} = \mathbf{I}_n$, the identity matrix of order n, then $\mathbf{u}_k \in \mathbf{R}^n$ and the ith component of $\mathbf{u}_k$ is applied to the ith component of the state vector. It is assumed that the initial condition x_o is specified. Let

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k, \tag{2.3}$$

where $\mathbf{z}_k \in \mathbf{R}^m$ for some positive integer m denotes the observation vector at time k, $\mathbf{h}$: $\mathbf{R}^n \rightarrow \mathbf{R}^m$ denotes the map (also known as the forward operator) that relates the model state $\mathbf{x}_k$ to the observation $\mathbf{z}_k$, and $\mathbf{v}_k$ is the observation noise vector, which is assumed to be white and Gaussian. That is, $\mathbf{v}_k \sim N(0, \mathbf{R}_k)$, where $\mathbf{R}_k \in \mathbf{R}^{m \times m}$ is a known positive definite matrix.

Define a performance measure

$$J = \sum_{k=0}^{N-1} V_k(\mathbf{x}_k, \mathbf{z}_k, \mathbf{u}_k), \tag{2.4}$$

where N is the number of observations, the cost functional V_k is a sum of two terms given by

$$V_k(\mathbf{x}_k, \mathbf{z}_k, \mathbf{u}_k) = V_k^0(\mathbf{x}_k, \mathbf{z}_k) + V_k^c(\mathbf{u}_k) \tag{2.5}$$

with

$$V_k^0(\mathbf{x}_k, \mathbf{z}_k) = \frac{1}{2}\langle [\mathbf{z}_k - \mathbf{h}(\mathbf{x}_k)], \quad \mathbf{R}^{-1}[\mathbf{z}_k - \mathbf{h}(\mathbf{x}_k)]\rangle, \tag{2.6}$$

$$V_k^c(\mathbf{u}_k) = \frac{1}{2}\langle \mathbf{u}_k, \mathbf{C}\mathbf{u}_k\rangle. \tag{2.7}$$

The notation $\langle a, b\rangle$ indicates the standard inner product, and $\mathbf{C} \in \mathbf{R}^{p\times p}$ is a given symmetric and positive definite matrix. Clearly V_k^0 denotes the energy in the normalized forecast error

$$e_k = \mathbf{z}_k - \mathbf{h}(\mathbf{x}_k) \tag{2.8}$$

and V_k^c accounts for the energy in the control input. The traditional variational methods use only V_k^0. For a given $\mathbf{R}^{-1}$, one can obtain a variety of tradeoffs between these two energy terms by appropriately choosing the matrix $\mathbf{C}$.

Define the Lagrangian L, obtained by augmenting the dynamical constraint in (2.1) with J in (2.4), as follows:

$$L = \sum_{k=0}^{N-1} \{V_k + \langle \boldsymbol{\lambda}_{k+1}, [\mathbf{M}(\mathbf{x}_k, \boldsymbol{\eta}_k, \mathbf{u}_k) - \mathbf{x}_{k+1}]\rangle\}, \tag{2.9}$$

where $\boldsymbol{\lambda}_k \in \mathbf{R}^n$ for $1 \le k \le N$ denotes the set of N undetermined Lagrangian multipliers or the costate variables. Now define the associated Hamiltonian function

$$H_k = H_k(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\eta}_k, \boldsymbol{\lambda}_{k+1}) = V_k + \langle \boldsymbol{\lambda}_{k+1}, \mathbf{M}(\mathbf{x}_k, \boldsymbol{\eta}_k, \mathbf{u}_k)\rangle. \tag{2.10}$$

Substituting (2.10) in (2.9), the latter becomes

$$L = \sum_{k=0}^{N-1} [H_k - \langle \boldsymbol{\lambda}_{k+1}, \mathbf{x}_{k+1}\rangle]. \tag{2.11}$$

By splitting the summation on the right-hand side of (2.11), we obtain

$$L = H_0 - \langle \boldsymbol{\lambda}_N, \mathbf{x}_N\rangle + \sum_{k=1}^{N-1} [H_k - \langle \boldsymbol{\lambda}_k, \mathbf{x}_k\rangle]. \tag{2.12}$$

Since $\boldsymbol{\eta}_k$ is specified, no variation of $\boldsymbol{\eta}_k$ is considered. Let δL be the induced increment in L resulting from the increments $\delta\mathbf{x}_k$ in $\mathbf{x}_k$ and $\delta\mathbf{u}_k$ in $\mathbf{u}_k$ for $0 \le k \le N-1$ and $\delta\boldsymbol{\lambda}_k$ in $\boldsymbol{\lambda}_k$ for $0 \le k \le N$. Since H_k is a scalar valued function of the vectors $\mathbf{x}_k$, $\mathbf{u}_k$, $\boldsymbol{\eta}_k$, and $\boldsymbol{\lambda}_{k+1}$, from the first principles (Lewis et al. 2006) we obtain

$$\begin{aligned}\delta L = {} & \langle \nabla_{\mathbf{x}_0} H_0, \delta\mathbf{x}_0\rangle + \langle \nabla_{\mathbf{u}_0} H_0, \delta\mathbf{u}_0\rangle - \langle \boldsymbol{\lambda}_N, \delta\mathbf{x}_N\rangle \\ & + \sum_{k=1}^{N} \langle \nabla_{\boldsymbol{\lambda}} H_{k-1} - \mathbf{x}_k, \delta\boldsymbol{\lambda}_k\rangle + \sum_{k=1}^{N-1} \langle \nabla_{\mathbf{x}} H_k - \boldsymbol{\lambda}_k, \delta\mathbf{x}_k\rangle \\ & + \sum_{k=1}^{N-1} \langle \nabla_{\mathbf{u}} H_k, \delta\mathbf{u}_k\rangle,\end{aligned} \tag{2.13}$$

where $\nabla_{\mathbf{x}} H_k \in \mathbf{R}^n$, $\nabla_{\mathbf{u}} H_k \in \mathbf{R}^p$, and $\nabla_{\boldsymbol{\lambda}} H_k \in \mathbf{R}^n$ are the gradients of H_k with respect to $\mathbf{x}_k$, $\mathbf{u}_k$, and $\boldsymbol{\lambda}_{k+1}$, respectively.

Recall that δL must be zero at the minimum, and in view of the arbitrariness of $\delta\mathbf{x}_k$, $\delta\mathbf{u}_k$, and $\delta\boldsymbol{\lambda}_k$, we readily obtain a set of necessary conditions expressed as follows, all for $0 \le k \le N-1$.

1) Condition 1: Model dynamics

The first summation, which is the fourth term on the right-hand side of (2.13), is zero when

$$\mathbf{x}_k = \nabla_{\boldsymbol{\lambda}} H_{k-1} = \frac{\partial H_{k-1}}{\partial \boldsymbol{\lambda}_k} \quad \text{for} \quad 1 \le k \le N-1. \tag{2.14}$$

Now computing the gradient of H_k in (2.10) with respect to $\boldsymbol{\lambda}_k$ and substituting it in (2.14), the latter becomes

$$\mathbf{x}_k = \overline{\mathbf{M}}(\mathbf{x}_{k-1}, \boldsymbol{\eta}_{k-1}) + \mathbf{B}\mathbf{u}_{k-1}, \tag{2.15}$$

which in fact turns out to be the model equations given in (2.2). Stated in other words, Pontryagin's method dictates that the sequence of states $\mathbf{x}_k$ arise as a solution of the model used as a strong constraint.

2) Condition 2: Costate or adjoint dynamics

The fifth summation on the right-hand side of (2.13) is zero when

$$\boldsymbol{\lambda}_k = \nabla_{\mathbf{x}} H_k = \frac{\partial H_k}{\partial \mathbf{x}_k} \quad \text{for} \quad 1 \le k \le N-1. \tag{2.16}$$

Computing the gradient of H_k in (2.10) with respect to the model state $\mathbf{x}_k$ and using it in (2.16), the latter becomes

$$\boldsymbol{\lambda}_k = \mathbf{D}_{\mathbf{x}_k}^{\mathrm{T}}(\mathbf{M})\boldsymbol{\lambda}_{k+1} + \nabla_{\mathbf{x}} V_k, \tag{2.17}$$

where $\nabla_{\mathbf{x}} V_k$ is the gradient of V_k in (2.5) given by

$$\nabla_{\mathbf{x}} V_k = \nabla_{\mathbf{x}} V_k^0 = \mathbf{D}_{\mathbf{x}}^{\mathrm{T}}(\mathbf{h})\mathbf{R}^{-1}[\mathbf{h}(\mathbf{x}_k) - \mathbf{z}_k], \tag{2.18}$$

which is the normalized forecast error viewed from the model space,

$$\mathbf{D}_{\mathbf{x}_k}(\mathbf{h}) = \left[\frac{\partial h_i}{\partial \mathbf{x}_j}\right]_{\mathbf{x}=\mathbf{x}_k} \in \mathbf{R}^{m\times n} \tag{2.19}$$

is the Jacobian of the forward operator $\mathbf{h}(\mathbf{x})$ with respect to the model state, and

$$\mathbf{D}_{\mathbf{x}_k}(\mathbf{M}) = \mathbf{D}_{\mathbf{x}_k}(\overline{\mathbf{M}}) = \left[\frac{\partial \overline{\mathbf{M}}_i}{\partial \mathbf{x}_j}\right]_{(\mathbf{x}=\mathbf{x}_k)} \in \mathbf{R}^{n\times n} \tag{2.20}$$

is the Jacobian of the model map $\mathbf{M}$ in (2.2).

Equation (2.17) is called the adjoint dynamics or the costate dynamics. By substituting (2.18) into (2.17), it takes a familiar form

$$\boldsymbol{\lambda}_k = \mathbf{D}^{\mathrm{T}}_{\mathbf{x}_k}(\mathbf{M})\boldsymbol{\lambda}_{k+1} + \mathbf{D}^{\mathrm{T}}_{\mathbf{x}_k}(\mathbf{h})\mathbf{R}^{-1}[\mathbf{h}(\mathbf{x}_k) - \mathbf{z}_k], \tag{2.21}$$

which is well known in the literature on 4DVAR methods (Lewis et al. 2006, 408–411).

3) CONDITION 3: STATIONARITY CONDITION

Similarly, combining the third summation, which is the sixth term with the second term on the right-hand side of (2.13), it follows that both of these two terms vanish when

$$0 = \nabla_{\mathbf{u}} H_k = \frac{\partial H_k}{\partial \mathbf{u}_k} \quad \text{for} \quad 0 \le k \le N. \tag{2.22}$$

Again computing the gradient of H_k in (2.10) with respect to the control $\mathbf{u}_k$ and using it in (2.22), the latter becomes

$$0 = \nabla_{\mathbf{u}} V_k + \mathbf{D}^{\mathrm{T}}_{\mathbf{u}}(\mathbf{M})\boldsymbol{\lambda}_{k+1}. \tag{2.23}$$

From (2.5) to (2.7) we get the gradient of V_k with respect to $\mathbf{u}_k$,

$$\nabla_{\mathbf{u}} V_k = \nabla_k V_k^c = \mathbf{C}\mathbf{u}_k, \tag{2.24}$$

and from (2.2) we get

$$\mathbf{D}_{\mathbf{u}}(\mathbf{M}) = \mathbf{B}, \tag{2.25}$$

which is the Jacobian of the model in (2.2) with respect to the control $\mathbf{u}_k$.

Now substituting (2.24) and (2.25) into (2.23), the structure of the optimal control is given by

$$\mathbf{u}_k = -\mathbf{C}^{-1}\mathbf{B}^{\mathrm{T}}\boldsymbol{\lambda}_{k+1}, \tag{2.26}$$

which is well defined since the matrix $\mathbf{C}$ in (2.7) is assumed to be a positive definite matrix. Notice that the second term on the right-hand side of (2.13) is already accounted for in (2.22). Thus, we are left with only the first and the third terms on the right-hand side of (2.13), which in turn provide the required boundary conditions.

Recall that $\mathbf{x}_0$ is given but $\mathbf{x}_N$ is free. Hence $\delta\mathbf{x}_0 = 0$ and $\delta\mathbf{x}_N$ is arbitrary. Thus, the first term on the right-hand side of (2.13) is automatically zero. The third term is zero by forcing

$$\boldsymbol{\lambda}_N = 0. \tag{2.27}$$

The above analysis naturally leads to a framework for optimal control, which is stated below.

(i) Step 1: Compute the optimal control

The structure of the optimal control sequence $\mathbf{u}_k$ is computed by solving the stationarity condition (2.22) and is given by (2.26).

(ii) Step 2: Solve the nonlinear TPBVP

Using the form of the optimal control in (2.26) in the model dynamics (2.15) or (2.2) and in the costate or adjoint dynamics in (2.21), we arrive at TPBVP given by

$$\mathbf{x}_{k+1} = \overline{\mathbf{M}}(\mathbf{x}_k, \boldsymbol{\eta}_k) - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^{\mathrm{T}}\boldsymbol{\lambda}_{k+1}, \tag{2.28}$$

$$\boldsymbol{\lambda}_k = \mathbf{D}^{\mathrm{T}}_{\mathbf{x}_k}(\mathbf{M})\boldsymbol{\lambda}_{k+1} + \mathbf{D}^{\mathrm{T}}_{\mathbf{x}_k}(\mathbf{h})\mathbf{R}^{-1}[\mathbf{h}(\mathbf{x}_k) - \mathbf{z}_k], \tag{2.29}$$

where the initial condition $\mathbf{x}_0$ for (2.28) is given and the final condition $\boldsymbol{\lambda}_N = 0$ is given for (2.29). Clearly, the solution (2.28) and (2.29) gives the optimal trajectory. A number of observations are in order.

The importance of the Hamiltonian formulation of the Euler–Lagrange necessary condition for the minimum stems from the simplicity and conciseness of the two first-order equations (2.14) and (2.16) involving the state and the costate/adjoint variables. This Hamiltonian formulation has been the basis of countless developments in physics (Goldstein 1980).

b. Computation of optimal control

Equation (2.28), a representation of the model dynamics, is solved forward in time starting from the known initial condition $\mathbf{x}_0$. But the adjoint (2.29), representing the costate/adjoint dynamics, is solved backward in time starting from $\boldsymbol{\lambda}_N = 0$. The two systems in (2.28) and (2.29) form a nonlinear coupled two-point boundary value problem, which in general does not admit to closed form solution. A number of numerical methods for solving (2.28) and (2.29) have been developed in the literature, a sampling of which is found in Roberts and Shipman (1972), Keller (1976), Polak (1997), and Bryson (1999). A closed form solution to the optimal control problem exists for the special case when the model

dynamics is linear and the cost function V_k is a quadratic form in state $\mathbf{x}_k$ and control $\mathbf{u}_k$. This special case is covered in section 3 of this paper.

c. Connection to 4DVAR

Consider the special case of an unforced model given by

$$\mathbf{x}_{k+1} = \overline{\mathbf{M}}(\mathbf{x}_k, \boldsymbol{\eta}_k), \tag{2.30}$$

where the initial condition $\mathbf{x}_0$ is arbitrary and

$$\overline{J} = \sum_{k=1}^{N-1} \overline{V}_k, \tag{2.31}$$

where $\overline{V}_k = \overline{V}_k^0(\mathbf{x}_k, \mathbf{z}_k) = V_k^0(\mathbf{x}_k, \mathbf{z}_k)$ is given by (2.6). Define the Lagrangian

$$\overline{L} = \overline{J} + \sum_{k=1}^{N-1} [\overline{H}_k(\mathbf{x}_k) - \langle \boldsymbol{\lambda}_k, \mathbf{x}_{k+1} \rangle], \tag{2.32}$$

where

$$\overline{H}_k = \overline{V}_k + \langle \overline{\boldsymbol{\lambda}}_{k+1}, \overline{\mathbf{M}}(\mathbf{x}_k, \boldsymbol{\eta}_k) \rangle. \tag{2.33}$$

By repeating the above argument we obtain the analog of (2.8) as

$$\delta L = \langle \nabla_{\mathbf{x}_0} \overline{H}_0, \delta \mathbf{x}_0 \rangle - \langle \overline{\boldsymbol{\lambda}}_N, \delta \mathbf{x}_N \rangle + \sum_{k=1}^{N} \langle (\nabla_{\boldsymbol{\lambda}} \overline{H}_{k-1} - \mathbf{x}_k), \delta \overline{\boldsymbol{\lambda}}_k \rangle + \sum_{k=1}^{N} \langle (\nabla_{\mathbf{x}} \overline{H}_k - \overline{\boldsymbol{\lambda}}_k), \delta \mathbf{x}_k \rangle. \tag{2.34}$$

The necessary conditions 1–3 for this special case take the following form.

1) CONDITION 1A: MODEL DYNAMICS

Vanishing of the third term on the right-hand side of (2.34) when $\delta \overline{\boldsymbol{\lambda}}_k$ is arbitrary leads to the condition

$$\mathbf{x}_k = \nabla_{\boldsymbol{\lambda}} \overline{H}_{k-1},$$

which in the light of (2.33) becomes the model equation

$$\mathbf{x}_k = \overline{\mathbf{M}}(\mathbf{x}_{k-1}, \boldsymbol{\eta}_{k-1}) \quad 1 \le k \le N. \tag{2.35}$$

2) CONDITION 2A: COSTATE/ADJOINT DYNAMICS

Since $\delta \mathbf{x}_k$ is arbitrary, vanishing of the fourth term on the right-hand side of (2.34) gives

$$\boldsymbol{\lambda}_k = \nabla_{\boldsymbol{\lambda}} \overline{H}_k,$$

which in the light of (2.33) becomes the costate dynamics given by

$$\boldsymbol{\lambda}_k = \mathbf{D}_{\mathbf{x}_k}^{\mathrm{T}}(\overline{\mathbf{M}}) \boldsymbol{\lambda}_{k+1} + \nabla_{\mathbf{x}} \overline{V}_k, \tag{2.36}$$

where $\nabla_{\mathbf{x}} \overline{V}_k = \nabla_{\mathbf{x}} V_k^0$ is given in (2.18).

Since $\mathbf{x}_N$ is free, $\delta \mathbf{x}_N$ is arbitrary. Hence, vanishing of the second term in (2.33) requires

$$\boldsymbol{\lambda}_N = 0. \tag{2.37}$$

Combining these, we readily see that (2.34) reduces to

$$\delta \overline{L} = \langle \nabla_{\mathbf{x}_k} \overline{H}_0, \delta \mathbf{x}_0 \rangle. \tag{2.38}$$

So, from first principles and using (2.33), it follows that

$$\frac{\partial L}{\partial \mathbf{x}_0} = \nabla_{\mathbf{x}_0} H_0 = \mathbf{D}_{\mathbf{x}_0}^{\mathrm{T}}(\overline{\mathbf{M}}) \boldsymbol{\lambda}_1 + \nabla_{\mathbf{x}_0} \overline{V}_k = \mathbf{D}_{\mathbf{x}_0}^{\mathrm{T}}(\overline{\mathbf{M}}) \boldsymbol{\lambda}_1 + \mathbf{D}_{\mathbf{x}_0}^{\mathrm{T}}(\mathbf{h}) \mathbf{R}^{-1} [\mathbf{h}(\mathbf{x}_0) - \mathbf{z}_0]. \tag{2.39}$$

The above development naturally leads to the standard 4DVAR algorithm (Lewis et al. 2006, 411–412), which is summarized below:

1) Starting from an arbitrary $\mathbf{x}_0$, compute the model solution $\mathbf{x}_k$ by iterating (2.35).
2) Using the observations $\mathbf{z}_k$, compute

$$f_k = \mathbf{D}_{\mathbf{x}_k}^{\mathrm{T}}(\mathbf{h}) \mathbf{R}^{-1} [\mathbf{h}(\mathbf{x}_k) - \mathbf{z}_k]. \tag{2.40}$$

3) Since $\nabla_{\mathbf{x}} \overline{V} = f_k$, using (2.40) in (2.36) iterate the adjoint dynamics backward to get the value of $\boldsymbol{\lambda}_1$.
4) Substitute $\boldsymbol{\lambda}_1$ in (2.39) to get the gradient $\partial L / \partial \mathbf{x}_0$.

It is easy to verify (Lewis et al. 2006, 386–389) that

$$\frac{\partial L}{\partial \mathbf{x}_0} = \frac{\partial \overline{J}}{\partial \mathbf{x}_0}, \tag{2.41}$$

where $\overline{J}$ is given by (2.31).

3. Optimal tracking: Linear dynamics

In this section we apply the minimum principle developed in section 2 to solve the problem of finding explicit form for optimal control or forcing that will drive the dynamics to track or follow the given set of observations when the model is linear and the performance measure is a quadratic function of the state and the control (Kalman 1963; Catlin 1989).

Let the deterministic dynamical model be given by

$$\mathbf{x}_{k+1} = \mathbf{M}\mathbf{x}_k + \boldsymbol{\eta}_k + \mathbf{B}\mathbf{u}_k, \tag{3.1}$$

where $\mathbf{M} \in \mathbf{R}^{n\times n}$, $\boldsymbol{\eta}_k \in \mathbf{R}^n$ is the intrinsic forcing term that is part of the model and $\mathbf{B} \in \mathbf{R}^{n\times p}$, which is the special case of (2.1). Let the observations be given by

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k, \tag{3.2}$$

where $\mathbf{H} \in \mathbf{R}^{m\times n}$ and $\mathbf{v}_k \sim N(0, \mathbf{R})$ and $\mathbf{R} \in \mathbf{R}^{m\times m}$ is the known positive definite matrix denoting the covariance of $\mathbf{v}_k$.

We consider the same cost functional given in (2.4)–(2.7). Substituting

$$\overline{\mathbf{M}}(\mathbf{x}, \boldsymbol{\eta}) = \mathbf{M}\mathbf{x} + \boldsymbol{\eta} \tag{3.3a}$$

and

$$\mathbf{M}(\mathbf{x}, \boldsymbol{\eta}, \mathbf{u}) = \mathbf{M}\mathbf{x} + \boldsymbol{\eta} + \mathbf{B}\mathbf{u} \tag{3.3b}$$

in the expression for the Lagrangian in (2.9) and in the subsequent developments in section 2, it can be verified that the necessary conditions for this linear case reduces to the following:

1) Structure of optimal control. From the stationarity condition developed in (2.22)–(2.26), it readily follows that the structure of the optimal control in this linear case is given by

$$\mathbf{u}_k = -\mathbf{C}^{-1}\mathbf{B}^{\mathrm{T}}\boldsymbol{\lambda}_{k+1}, \tag{3.4}$$

which is the same as in the nonlinear case treated in section 2.

2) The linear two-point boundary value problem. Substituting the special form of the dynamics and the observation given by (3.1)–(3.3) and the expression for $\mathbf{u}_k$ given by (3.4) in (2.28) and (2.29), the latter pair of equations become

$$\mathbf{x}_{k+1} = \mathbf{M}\mathbf{x}_k + \boldsymbol{\eta}_k - (\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^{\mathrm{T}})\boldsymbol{\lambda}_{k+1}, \tag{3.5}$$

$$\boldsymbol{\lambda}_k = \mathbf{M}^{\mathrm{T}}\boldsymbol{\lambda}_{k+1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}_k - \mathbf{z}_k), \tag{3.6}$$

where we have used the fact that $\mathbf{h}(\mathbf{x}) = \mathbf{H}\mathbf{x}$ and $\mathbf{D}_{\mathbf{x}}(\mathbf{h}) = \mathbf{H}$. The initial condition for (3.5) is the given $\mathbf{x}_0$ and the final condition for (3.6) is $\boldsymbol{\lambda}_N = 0$. Again, recall that (3.6) is the well-known adjoint equation that routinely arises in the 4DVAR analysis (Lewis et al. 2006, 408–412). For later reference we rewrite (3.5) and (3.6) as

$$\begin{bmatrix} \mathbf{x}_{k+1} \\ \boldsymbol{\lambda}_k \end{bmatrix} = \begin{bmatrix} \mathbf{M} & -\mathbf{E} \\ \mathbf{F} & \mathbf{M}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_k \\ \boldsymbol{\lambda}_{k+1} \end{bmatrix} + \begin{bmatrix} 0 \\ -\mathbf{W} \end{bmatrix} \mathbf{z}_k, \tag{3.7}$$

where $\mathbf{E} = \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^{\mathrm{T}}$, $\mathbf{F} = \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$, and $\mathbf{W} = \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}$.

It turns out this special linear TPBVP can be transformed into a pair of initial value problems using the sweep method, which in turn can be easily solved. By exploiting the structure of (3.5) and (3.6), it can be verified (see appendix A for details) that $\boldsymbol{\lambda}_k$ is an affine function of the state $\mathbf{x}_k$.

Consequently, we posit that $\boldsymbol{\lambda}_k$ is related to $\mathbf{x}_k$ via a general affine transformation of the type

$$\boldsymbol{\lambda}_k = \mathbf{P}_k\mathbf{x}_k - \mathbf{g}_k. \tag{3.8}$$

Substituting (3.8) in the state equation in (3.7) and simplifying, we get

$$\mathbf{x}_{k+1} = (\mathbf{I} + \mathbf{E}\mathbf{P}_{k+1})^{-1}(\mathbf{M}\mathbf{x}_k + \mathbf{E}\mathbf{g}_{k+1}). \tag{3.9}$$

Again substituting (3.8) and (3.9) in the costate equation in (3.7), after simplifying we get

$$\begin{aligned}[\mathbf{g}_k + \mathbf{M}^{\mathrm{T}}\mathbf{P}_{k+1}(\mathbf{I} + \mathbf{E}\mathbf{P}_{k+1})^{-1}(\boldsymbol{\eta}_k + \mathbf{E}\mathbf{g}_{k+1}) - \mathbf{M}^{\mathrm{T}}\mathbf{g}_{k+1} \\ - \mathbf{W}\mathbf{z}_k] + [-\mathbf{P}_k + \mathbf{M}^{\mathrm{T}}\mathbf{P}_{k+1}(\mathbf{I} + \mathbf{E}\mathbf{P}_{k+1})^{-1}\mathbf{M} + \mathbf{F}]\mathbf{x}_k = 0.\end{aligned} \tag{3.10}$$

Since (3.10) must hold good for all $\mathbf{x}_k$, we immediately obtain equations that define the evolution of the matrix $\mathbf{P}_k$ and the vector $\mathbf{g}_k$ as follows:

$$\mathbf{P}_k = \mathbf{M}^{\mathrm{T}}\mathbf{P}_{k+1}(\mathbf{I} + \mathbf{E}\mathbf{P}_{k+1})^{-1}\mathbf{M} + \mathbf{F}, \tag{3.11}$$

which is a nonlinear matrix Riccati equation and

$$\begin{aligned}\mathbf{g}_k = \mathbf{M}^{\mathrm{T}}\mathbf{g}_{k+1} - \mathbf{M}^{\mathrm{T}}\mathbf{P}_{k+1}(\mathbf{I} + \mathbf{E}\mathbf{P}_{k+1})^{-1} \\ \times (\boldsymbol{\eta}_k + \mathbf{E}\mathbf{g}_{k+1}) + \mathbf{W}\mathbf{z}_k.\end{aligned} \tag{3.12}$$

Since $\boldsymbol{\lambda}_N = 0$ and $\mathbf{x}_N$ is arbitrary, from (3.8) it is immediately clear that

$$\mathbf{P}_N = 0 \quad \text{and} \quad \mathbf{g}_N = 0. \tag{3.13}$$

Against this backdrop, we now state the algorithm for computing the optimal control and the optimal trajectory.

- Step 1
  Given (3.1)–(3.3), compute $\mathbf{E} = \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^{\mathrm{T}}$, $\mathbf{F} = \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$, and $\mathbf{W} = \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}$. Solve the nonlinear matrix Ricatti difference equation in (3.11) for $\mathbf{P}_k$ backward starting at $\mathbf{P}_N = 0$. Since this computation is independent of the observations, it can be precomputed and stored if needed.
- Step 2
  Solve the linear vector difference equation in (3.12) for $\mathbf{g}_k$ backward in time starting from $\mathbf{g}_N = 0$. Notice that $\mathbf{g}_k$ depends on the observations and the

intrinsic forcing $\boldsymbol{\eta}_k$ that is part of the given model. It will be seen that the impact of the observations on the optimal control is through $\mathbf{g}_k$.

- Step 3

  Once $\mathbf{P}_k$ and $\mathbf{g}_k$ are known, we can compose the optimal control using (3.4) and (3.8):

$$\mathbf{u}_k = -\mathbf{C}^{-1}\mathbf{B}^{\mathrm{T}}(\mathbf{P}_{k+1}\mathbf{x}_{k+1} - \mathbf{g}_{k+1}). \tag{3.14}$$

  Using (3.1) in (3.14), the latter becomes

$$\mathbf{u}_k = -\mathbf{C}^{-1}\mathbf{B}^{\mathrm{T}}[\mathbf{P}_{k+1}(\mathbf{M}\mathbf{x}_k + \boldsymbol{\eta}_k + \mathbf{B}\mathbf{u}_k) - \mathbf{g}_{k+1}]. \tag{3.15}$$

  Premultiplying both sides by C and simplifying, we get an explicit expression for the optimal control as

$$\mathbf{u}_k = -\mathbf{K}_k\mathbf{x}_k + \mathbf{G}_k\mathbf{g}_{k+1} - \mathbf{K}_k\mathbf{M}^{-1}\boldsymbol{\eta}_k, \tag{3.16}$$

  where the feedback gain $\mathbf{K}_k$ is given by

$$\mathbf{K}_k = (\mathbf{C} + \mathbf{B}^{\mathrm{T}}\mathbf{P}_{k+1}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P}_{k+1}\mathbf{M} \tag{3.17}$$

  and the feedforward gain $\mathbf{G}_k$ is given by

$$\mathbf{G}_k = (\mathbf{C} + \mathbf{B}^{\mathrm{T}}\mathbf{P}_{k+1}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}. \tag{3.18}$$

  From (3.1) and (3.16), the optimal trajectory is then given by

$$\mathbf{x}_{k+1} = (\mathbf{M} - \mathbf{B}\mathbf{K}_k)\mathbf{x}_k + \mathbf{B}\mathbf{G}_k\mathbf{g}_{k+1} + (\mathbf{I} - \mathbf{B}\mathbf{K}_k\mathbf{M}^{-1})\boldsymbol{\eta}_k \tag{3.19}$$

  or as

$$\mathbf{x}_{k+1} = \mathbf{M}\mathbf{x}_k - \mathbf{B}(\mathbf{K}_k\mathbf{x}_k - \mathbf{G}_k\mathbf{g}_{k+1}) + (\mathbf{I} - \mathbf{B}\mathbf{K}_k\mathbf{M}^{-1})\boldsymbol{\eta}_k. \tag{3.20}$$

The second term on the right-hand side of (3.20) is indeed the optimal forcing term $\mathbf{B}\mathbf{u}_k$ and it plays a dual role. First, it forces the model trajectory toward the observations where the measure of closeness depends on the choice of p, the dimension of the control vector $\mathbf{u}_k$, the matrix $\mathbf{B} \in \mathbf{R}^{n\times p}$, and the matrix $\mathbf{C} \in \mathbf{R}^{p\times p}$, where it is assumed that the observational error covariance matrix $\mathbf{R}$ is given. Consequently, $\mathbf{B}\mathbf{u}_k$ contains information about the model error. Thus, for a given value of $\mathbf{R}$ and a prespecified measure of closeness between the observations and the model trajectory, one can, in principle, obtain a family of optimal control $\mathbf{u}_k$ to achieve this goal by suitably varying the integer p, $(1 \le p \le n)$, and $\mathbf{B} \in \mathbf{R}^{n\times p}$ and $\mathbf{C} \in \mathbf{R}^{p\times p}$ with $\mathbf{C}$ being symmetric and positive definite.

4. Dynamical constraint: Linear Burgers's equation

To illustrate Pontryagin's method, we choose a dynamic constraint that follows the theme of Platzman's classical study of Burgers's equation (Platzman 1964). In that study, Platzman investigated the evolution of an initial single primary sine wave over the interval $[0, 2\pi]$. The governing dynamics described the transfer of energy from this primary wave to waves of higher wavenumber as the wave neared the breaking point. In a tour de force with spectral dynamics, Platzman obtained a closed form solution for the Fourier amplitudes and then analyzed the consequences of truncated spectral expansions. The contribution was instrumental in helping dynamic meteorologists understand the penalties associated with truncated spectral weather forecasting in the early days of numerical weather prediction.

We maintain the spirit of Platzman's investigation but in a somewhat simplified form. Whereas the nonlinear dynamic law advects the wave with the full spectrum of Fourier components, we choose to advect with only the initial primary wave—$\sin(x)$. This problem retains the transfer of energy from the primary wave to the higher wavenumber components as the wave steepens, but the more complex phenomenon of folding over or breaking of the wave is absent in this linear problem.

a. Model and its analytic solution

The governing dynamics for the linear Burgers's equation is

$$\mathbf{q}_t + \sin(x) \times \mathbf{q}_x = 0, \quad 0 \le x \le 2\pi, \tag{4.1}$$

with initial condition $\mathbf{q}(x, 0) = \sin(x)$ and boundary conditions $\mathbf{q}(0, t) = \mathbf{q}(2\pi, t) = 0$. Following Platzman (1964), we seek a solution to (4.1) by the method of characteristics (Carrier and Pearson 1976). The characteristics of (4.1) are given by

$$\frac{1 + \cos(x)}{1 - \cos(x)}e^{2t} = \frac{1 + \cos(x_0)}{1 - \cos(x_0)}, \tag{4.2}$$

where x_0 is the intersection of a particular characteristic curve with the line of initial time ($t = 0$). Using the mathematical expression for the characteristics in concert with the initial condition, the analytic solution is found to be

$$\mathbf{q}(x, t) = \frac{2e^t \sin(x)}{1 + e^{2t} + \cos(x) \times (e^{2t} - 1)}. \tag{4.3}$$

From this analytic solution, the profiles of the wave at times $t = 0, 0.5, 1.0, 1.5$, and 2.0 are shown in Fig. 1. The slope of the wave is finite at $x = \pi$ but approaches ∞ as $t \to \infty$.

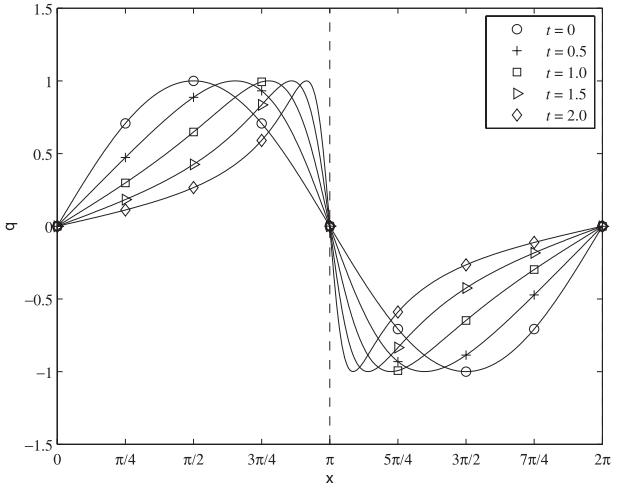

FIG. 1. A plot of the solution $\mathbf{q}(\mathbf{x}, t)$ in (4.3) at times $t = 0$, 0.5, 1, 1.5, and 2.

Let $\tilde{\mathbf{q}}_k(t)$ be the (exact) value of the kth Fourier coefficient of the solution $\mathbf{q}(x, t)$ in (4.3) given by

$$\tilde{\mathbf{q}}_k(t) = \frac{1}{\pi}\int_0^{2\pi} \mathbf{q}(x,t)\sin(kx)\,dx. \tag{4.4}$$

Define the vector

$$\tilde{\mathbf{Q}}_t = [\tilde{\mathbf{q}}_1(t), \tilde{\mathbf{q}}_2(t), \ldots, \tilde{\mathbf{q}}_n(t)]^{\mathrm{T}} \in \mathbf{R}^n \tag{4.5}$$

of the first n Fourier coefficients of $\mathbf{q}(x, t)$. The values of the coefficients $\tilde{\mathbf{q}}_k(t)$ (computed using the well-known quadrature formula) for $1 \le k \le n = 8$ and $0 \le t \le 2.0$ in steps of $\Delta t = 0.2$ are given in (rows of) Table 1.

An n-mode approximation [resulting from spectral truncation to $\mathbf{q}(x, t)$] is then given by

$$\tilde{\mathbf{q}}(x,t) = \sum_{k=1}^{n} \tilde{\mathbf{q}}_k(t)\sin(kx). \tag{4.6}$$

A comparison of the exact solution $\mathbf{q}(x, t)$ with the four-mode approximation $\tilde{\mathbf{q}}_1(x,t)$ and the eight-mode approximation $\tilde{\mathbf{q}}_2(x,t)$ obtained from (4.6) with $n = 4$ and 8, respectively, at $t = 2.0$ is given in Fig. 2. As to be expected, the eight-mode approximation is closer to the true solution than is the four-mode approximation. Further, the errors are the greatest at the point of extreme steepness of waves.

b. The low-order model

In demonstrating the power of Pontryagin's method developed in sections 2 and 3, our immediate goal is to obtain a discrete time model representative of (3.1). There are at least two ways, in principle, to achieve this goal. The first way is to directly discretize (4.1) by embedding a grid in the two-dimensional domain with $0 \le x \le 2\pi$ and $t \ge 0$. Second is to project the infinite dimensional system in (4.1) onto a finite dimensional space using the standard Galerkin projection method and obtain a system of n ordinary differential equations (ODEs) describing the evolution of the Fourier amplitudes $\mathbf{q}_i(t)$ in (4.4), $1 \le i \le n$. The resulting nth-order system is known as the low-order model (LOM). Then LOM can be discretized using one of several known methods. In this paper we embrace this latter approach using LOM.

TABLE 1. Values of the first eight Fourier coefficients of $\mathbf{q}(\mathbf{x}, t)$ in (4.3) for various times computed using quadrature formula.

t	$\tilde{\mathbf{q}}_1(t)$	$\tilde{\mathbf{q}}_2(t)$	$\tilde{\mathbf{q}}_3(t)$	$\tilde{\mathbf{q}}_4(t)$	$\tilde{\mathbf{q}}_5(t)$	$\tilde{\mathbf{q}}_6(t)$	$\tilde{\mathbf{q}}_7(t)$	$\tilde{\mathbf{q}}_8(t)$
0.0	1	0	0	0	0	0	0	0
0.2	0.990	−0.099	0.010	−0.001	0.000	−0.000	0.000	−0.000
0.4	0.961	−0.190	0.037	−0.007	0.001	−0.000	0.000	−0.000
0.6	0.915	−0.267	0.078	−0.023	0.007	−0.002	0.001	−0.000
0.8	0.856	−0.325	0.124	−0.047	0.018	−0.007	0.003	−0.001
1.0	0.786	−0.363	0.168	−0.078	0.036	−0.017	0.008	−0.004
1.2	0.712	−0.382	0.205	−0.110	0.059	−0.032	0.017	−0.009
1.4	0.634	−0.384	0.232	−0.140	0.085	−0.051	0.031	−0.019
1.6	0.559	−0.371	0.247	−0.164	0.109	−0.072	0.048	−0.032
1.8	0.487	−0.349	0.250	−0.179	0.128	−0.092	0.066	−0.047
2.0	0.420	−0.320	0.244	−0.153	0.141	−0.108	0.082	−0.062

The Fourier series of $\mathbf{q}(x, t)$ consists of an infinite series of sine waves given by

$$\mathbf{q}(x,t) = \sum_{n=1}^{\infty} \mathbf{q}_n(t)\sin(nx). \tag{4.7}$$

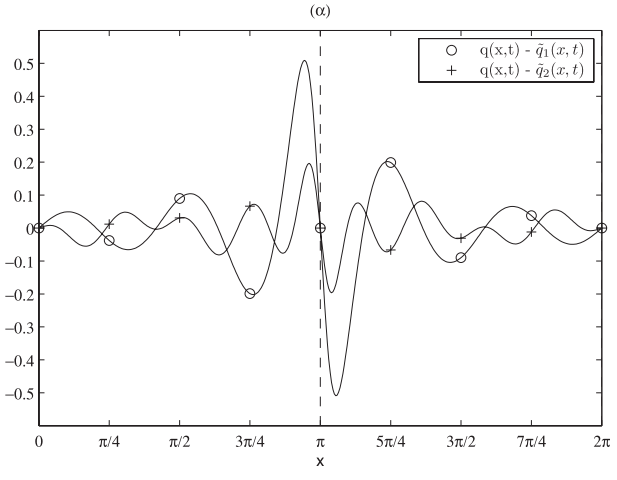

FIG. 2. Comparison of the error $\mathbf{q}(\mathbf{x}, t) - \tilde{\mathbf{q}}_1(\mathbf{x}, t)$ in the four-mode approximation $\tilde{\mathbf{q}}_1(\mathbf{x}, t) = \sum_{i=1}^{4} \tilde{\mathbf{q}}_i(t)\sin(i\mathbf{x})$ and the error $\mathbf{q}(\mathbf{x}, t) - \tilde{\mathbf{q}}_2(\mathbf{x}, t)$ in the eight-mode approximation $\tilde{\mathbf{q}}_2(\mathbf{x}, t) = \sum_{i=1}^{8} \tilde{\mathbf{q}}_i(t)\sin(i\mathbf{x})$ at $t = 2.0$. Fourier coefficients $\tilde{\mathbf{q}}_i(t)$ at $t = 2$ are given in Table 1.

An LOM(n) describing evolution of the amplitudes of the spectral components are obtained by exploiting the orthogonality properties of the $\sin(ix)$ functions for $1 \le i \le n$. Substituting (4.4) into (4.1), multiplying both sides by $\sin(ix)$, and integrating both sides from 0 to 2π, we obtain the LOM(n) (also known as the spectral model):

$$\frac{d\mathbf{q}(t)}{dt} = \mathbf{A}\mathbf{q}(t), \tag{4.8}$$

where

$$\mathbf{q}(t) = [\mathbf{q}_1(t), \mathbf{q}_2(t), \ldots, \mathbf{q}_n(t)]^{\mathrm{T}}, \quad \mathbf{q}(0) = (1, 0, 0, \ldots, 0)^{\mathrm{T}} \tag{4.9}$$

as its initial condition and the matrix $\mathbf{A}$ given by

$$\mathbf{A} = \frac{1}{2}\begin{bmatrix} 0 & c_1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ a_2 & 0 & c_2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & a_3 & 0 & c_3 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & a_{n-1} & 0 & c_{n-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & a_n & 0 \end{bmatrix}, \tag{4.10}$$

where $a_i = -(i-1), c_i = (i+1)$. An example for $n = 4$ is given by

$$\mathbf{A} = \frac{1}{2}\begin{bmatrix} 0 & 2 & 0 & 0 \\ -1 & 0 & 3 & 0 \\ 0 & -2 & 0 & 4 \\ 0 & 0 & -3 & 0 \end{bmatrix}. \tag{4.11}$$

We now state a number of interesting properties of the solution of the LOM(n) in (4.8).

1) CONSERVATION OF ENERGY

Consider a quadratic form $E(\mathbf{q})$ representing generalized energy and given by

$$E(\mathbf{q}) = \frac{1}{2}\mathbf{q}^{\mathrm{T}}\mathbf{K}\mathbf{q} = \frac{1}{2}\sum_{k=1}^{n}\mathbf{K}_k\mathbf{q}_k^2, \tag{4.12}$$

where

$$\mathbf{K} = \mathrm{Diag}(1, 2, 3, \ldots, i, \ldots, n) \tag{4.13}$$

is a diagonal matrix with the indicated entries as its diagonal elements. It can be verified that the time derivative of $E(\mathbf{q})$ evaluated along the solution of (4.8) is given by

$$\frac{dE(\mathbf{q})}{dt} = \mathbf{q}^{\mathrm{T}}\mathbf{K}\frac{d\mathbf{q}}{dt} = \mathbf{q}^{\mathrm{T}}\mathbf{K}\mathbf{A}\mathbf{q}. \tag{4.14}$$

From the form of $\mathbf{K}$ in (4.13) and $\mathbf{A}$ in (4.10), it is an easy exercise to verify that the product $\mathbf{KA}$ is a skew-symmetric matrix given by

$$\mathbf{KA} = \frac{1}{2}\begin{bmatrix} 0 & s_1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -s_1 & 0 & s_2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -s_2 & 0 & s_3 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -s_{n-2} & 0 & s_{n-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & -s_{n-1} & 0 \end{bmatrix}, \tag{4.15}$$

where $s_i = i(i+1)$ for $1 \le i \le n-1$. Hence, it can be verified that the quadratic form $\mathbf{q}^{\mathrm{T}}\mathbf{KA}q$ is zero, which in turn implies that the energy $E(\mathbf{q})$ is conserved by (4.8); that is,

$$\frac{dE(\mathbf{q})}{dt} = 0. \tag{4.16}$$

An immediate consequence of (4.16) is that the solution $\mathbf{q}(t)$ of (4.8) always lies on the surface of an n-dimensional ellipsoid defined by

$$\sum_{k=1}^{n}\mathbf{K}_k\mathbf{q}_k^2(t) = \sum_{k=1}^{n}\mathbf{K}_k\mathbf{q}_k^2(0) = 1. \tag{4.17}$$

Clearly, the length of the kth semiaxis of this ellipsoid is given by $(1/k)^{1/2}$. Hence the volume of this ellipsoid is given by

$$\text{Volume} = \frac{4}{3}\pi\left(\frac{1}{n!}\right)^{1/2}. \tag{4.18}$$

Since $n! = O(2^{n\log n})$, it turns out that the volume of this ellipsoid goes to zero at an exponential rate as n increases signaling degeneracy for large n.

2) SOLUTION OF LOM(n) IN (4.8)

Much like the PDE (4.1), its LOM(n) counterpart in (4.8) can also be solved exactly. The process of obtaining its solution is quite involved. To minimize the digression from the main development, we have chosen to describe this solution process in appendix B. The eigenstructure of $\mathbf{A}$, its Jordan canonical form, and the form of the general solution of (4.6) are discussed in detail in appendix B. Our goal is to use the exact solution of (4.8) given in appendix B to calibrate the choice of Δt—the time discretization interval to be used in the following section.

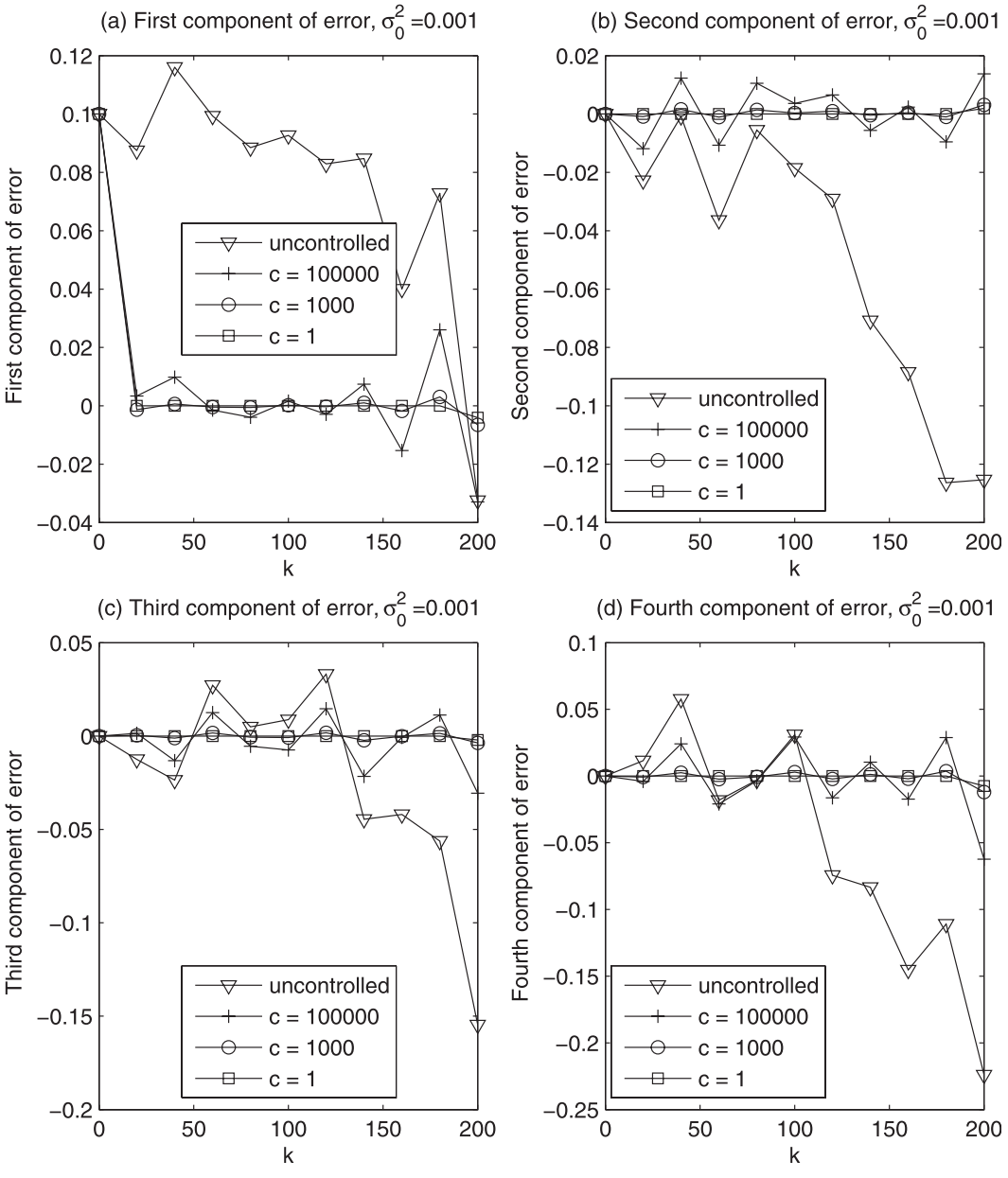

FIG. 3. Comparison of the four components of the uncontrolled error $e_0 = \boldsymbol{\xi}_k - \mathbf{z}_k$ and the controlled error $e_c = \mathbf{x}_k - \mathbf{z}_k$ for $p = 4$, $\mathbf{B} = \mathbf{I}_4$, $c = \{100\,000, 1000, 1\}$.

5. Numerical experiments

Discretizing the spectral model in (4.8) with $n = 4$ using the first-order Euler scheme, we obtain

$$\boldsymbol{\xi}_{k+1} = \mathbf{M}\boldsymbol{\xi}_k, \tag{5.1}$$

where $\boldsymbol{\xi}_k = \mathbf{q}(t = k\Delta t)$ and Δt denotes the length of the time interval used in time discretization and

$$\mathbf{M} = (\mathbf{I} + \Delta t\mathbf{A}) \in \mathbf{R}^{n\times n}, \tag{5.2}$$

where $\mathbf{A} \in \mathbf{R}^{4\times 4}$ is given in (4.11) and the initial condition in (4.9).

Pontryagin's approach requires the addition of the forcing term to (5.1). The forced version of (5.1) is then represented as

$$\mathbf{x}_{k+1} = \mathbf{M}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k, \tag{5.3}$$

where $\mathbf{u}_k \in \mathbf{R}^p$ and $\mathbf{B} \in \mathbf{R}^{n\times p}$. Clearly (5.3) is the same as (3.1) with $\boldsymbol{\eta}_k \equiv 0$ and $\mathbf{x}_0 = \mathbf{q}(0)$.

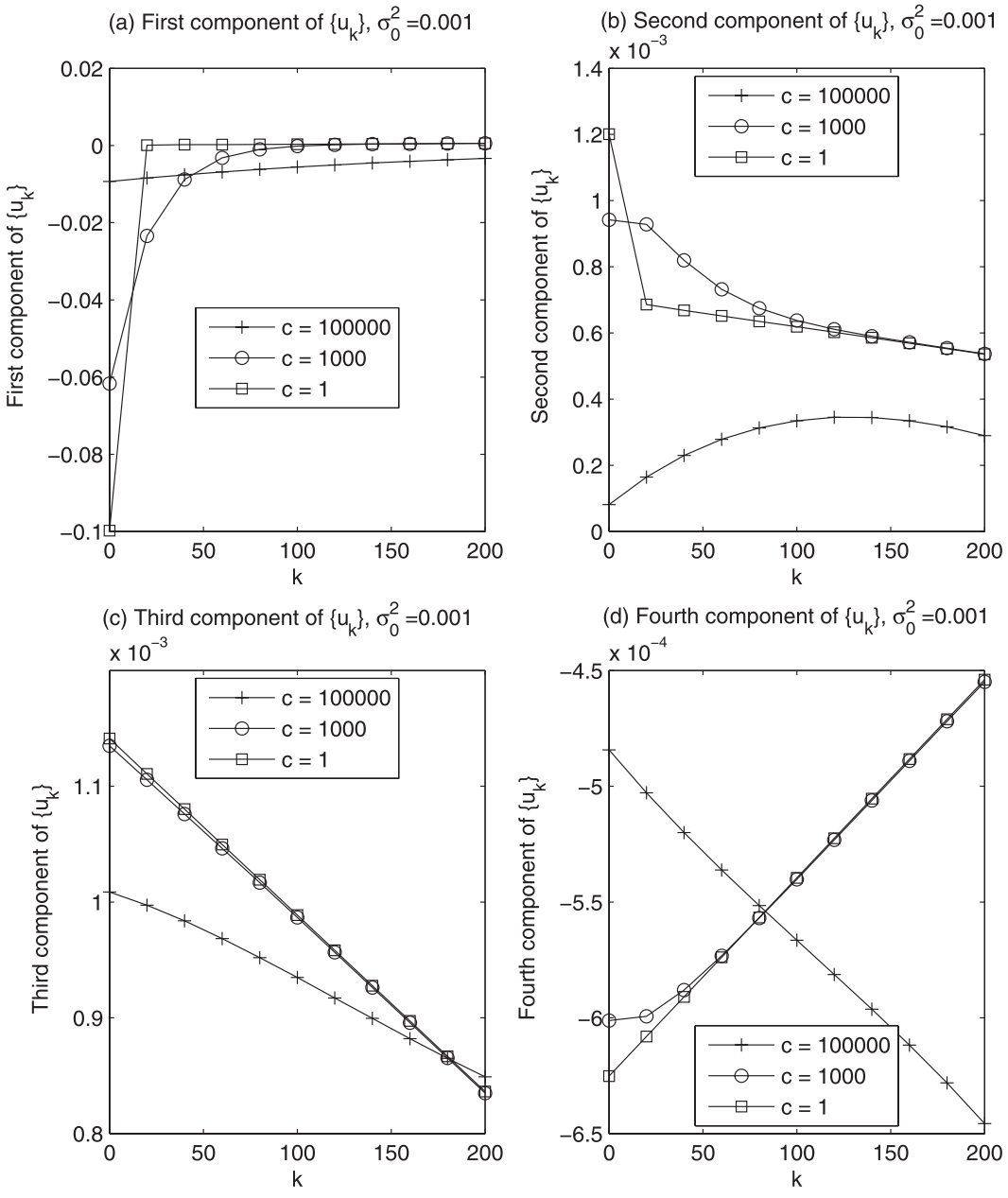

FIG. 4. Comparison of four components of the control sequence $\{\mathbf{u}_k\}$ for $p = 4$, $\mathbf{B} = \mathbf{I}_4$, and $c = \{100\,000, 1000, 1\}$.

Equation (5.3) defines the evolution of the spectral amplitudes. Compared to the original equation, the spectral model in (5.3) has two types of model errors: one from the spectral truncation in the Galerkin projection and one due to finite differencing in (4.8) using the first-order method.

Observations

We propose to use the exact Fourier coefficient vector $\tilde{\mathbf{Q}}_k = \tilde{\mathbf{Q}}_t$ at $t = k\Delta t$ in (4.5) corrupted by additive noise as the observations in our numerical experiments. Define

$$\mathbf{z}_k = \tilde{\mathbf{Q}}_k + \boldsymbol{\nu}_k, \tag{5.4}$$

where $\mathbf{z}_k \in \mathbf{R}^n$, $\tilde{\mathbf{Q}}_k \in \mathbf{R}^n$, $\boldsymbol{\nu}_k \sim N(0, \mathrm{R})$, and $\mathbf{R} = \sigma_0^2 \mathbf{I}_n$.

Comparing (5.4) with (3.2), it is immediate that $m = n$ and $\mathbf{H} = \mathbf{I}_n$.

The form of the functional V_k is given by

$$V_k = \frac{1}{2}(\mathbf{z}_k - \mathbf{x}_k)^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{z}_k - \mathbf{x}_k) + \frac{1}{2} \mathbf{u}_k^{\mathrm{T}} \mathbf{C} \mathbf{u}_k, \tag{5.5}$$

where $\mathbf{C} \in \mathbf{R}^{p \times p}$ is a symmetric and positive definite matrix.

Applying the results from section 3, it follows that

$$\mathbf{u}_k = -\mathbf{C}^{-1}\mathbf{B}^{\mathrm{T}}\boldsymbol{\lambda}_{k+1}, \tag{5.6}$$

where

$$\boldsymbol{\lambda}_k = \mathbf{M}^{\mathrm{T}}\boldsymbol{\lambda}_{k+1} + \mathbf{R}^{-1}(\mathbf{x}_k - \mathbf{z}_k) \tag{5.7}$$

with $\boldsymbol{\lambda}_N = 0$.

The TPBVP problem in (5.3) and (5.7) is then solved using the sweep method described in section 3. Accordingly,

$$\boldsymbol{\lambda}_k = \mathbf{P}_k\mathbf{x}_k - \mathbf{g}_k, \tag{5.8}$$

where

$$\mathbf{P}_k = \mathbf{M}^{\mathrm{T}}\mathbf{P}_{k+1}(\mathbf{I} + \mathbf{E}\mathbf{P}_{k+1})^{-1}\mathbf{M} + \mathbf{F}, \tag{5.9}$$

$$\mathbf{g}_k = \mathbf{M}^{\mathrm{T}}\mathbf{g}_{k+1} - \mathbf{M}^{\mathrm{T}}\mathbf{P}_{k+1}(\mathbf{I} + \mathbf{E}\mathbf{P}_{k+1})^{-1}\mathbf{E}\mathbf{g}_{k+1} + \mathbf{W}\mathbf{z}_k, \tag{5.10}$$

where $\mathbf{E} = \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^{\mathrm{T}}$, $\mathbf{F} = \mathbf{R}^{-1}$, $\mathbf{W} = \mathbf{R}^{-1}$, $\mathbf{P}_N = 0$, and $\mathbf{g}_N = 0$.

Solving (5.9) and (5.10), we then assemble $\mathbf{u}_k$ using (3.14)–(3.18). Substituting it in (5.1) we get the optimal solution.

1) EXPERIMENT 1

In this first experiment, we set $n = m = p = 4$, $\mathbf{B} = \mathbf{I}_4$, and $\mathbf{u}_k \in \mathbf{R}^4$. The uncontrolled model is

$$\boldsymbol{\xi}_{k+1} = \mathbf{M}\boldsymbol{\xi}_k \tag{5.11}$$

and the controlled model is

$$\mathbf{X}_{k+1} = \mathbf{M}\mathbf{X}_k + \mathbf{B}\mathbf{u}_k, \tag{5.12}$$

with $\mathbf{M} = (\mathbf{I} + \Delta t\mathbf{A})$ and $\mathbf{A}$ is given in (4.11).

Both models start from the same initial condition $\boldsymbol{\xi}_0 = \mathbf{x}_0 = (1.1, 0, 0, 0)^{\mathrm{T}}$, which is different from the one that was used to generate the observations. Consequently, the solution to the unforced model in (5.11) inherits three types of errors: the first because of the spectral truncation, the second because of finite differencing, and the third owing to error in the initial condition. The power of the Pontryagin's approach is to compute the optimal control $\mathbf{u}_k$ such that the term $\mathbf{B}\mathbf{u}_k$ compensates for all the three types of errors.

The observation vector $\mathbf{z}_k \in \mathbf{R}^4$ is given by

$$\mathbf{z}_k = \tilde{\mathbf{Q}}_k + \boldsymbol{\nu}_k \tag{5.13}$$

for $1 \le k \le 10$, where $\tilde{\mathbf{Q}}_k = [\tilde{\mathbf{q}}_1(k), \tilde{\mathbf{q}}_2(k), \tilde{\mathbf{q}}_3(k), \tilde{\mathbf{q}}_4(k)]^{\mathrm{T}}$ given in Table 1, and $\boldsymbol{\nu}_k \sim N(0, \mathbf{R})$.

TABLE 2. Root-mean-square errors of the controlled and uncontrolled model solution with observations ($p = 4$, $\mathbf{B} = \mathbf{I}_4$).

σ_0^2	c	RMS$_{e1}$	RMS$_{e2}$	RMS$_{e3}$	RMS$_{e4}$
Uncontrolled		0.0850	0.0658	0.0551	0.0958
0.001	100 000	0.0333	0.0090	0.0140	0.0258
	1000	0.0302	0.0013	0.0016	0.0042
	1	0.0302	0.0006	0.0006	0.0023
0.005	100 000	0.0654	0.0438	0.0345	0.0465
	1000	0.0323	0.0084	0.0061	0.0112
	1	0.0302	0.0005	0.0007	0.0027
0.01	100 000	0.0901	0.1160	0.1200	0.1010
	1000	0.0368	0.0337	0.0360	0.0284
	1	0.0302	0.0005	0.0072	0.0028

It is further assumed that $\mathbf{R} = \sigma_0^2\mathbf{I}_n$ and $\mathbf{C} = c\mathbf{I}_p$. Substituting these in the expression for V_k in (2.5)–(2.7), it can be verified that

$$V_k = \sigma_0^2\langle \mathbf{z}_k - \mathbf{H}\mathbf{x}_k, \mathbf{z}_k - \mathbf{H}\mathbf{x}_k\rangle + c\langle \mathbf{u}_k, \mathbf{u}_k\rangle. \tag{5.14}$$

A comparison of the evolution of the four components of the uncontrolled error, $e_0 = \boldsymbol{\xi}_k - \mathbf{z}_k \in \mathbf{R}^4$, and the corresponding components of the controlled error, $e_c = \boldsymbol{\xi}_k - \mathbf{z}_k \in \mathbf{R}^4$, when $\sigma^2 = 0.001$ fixed but c is varied through 10^5, 10^3, and 1, are given in Figs. 3a–d. It is clear that the magnitudes of the individual components of the controlled error are uniformly (in time k) less than the magnitudes of the corresponding components of the uncontrolled error. Further, the magnitudes of the controlled error decrease with the decrease in the value of the control parameter c.

This behavior can be easily explained using (5.14). When the value of the control parameter c is large (for a fixed $\mathbf{R}^{-1}$), the minimization process forces $\mathbf{u}_k$ to be small. However, if c is small, the minimization allows for larger value of $\mathbf{u}_k$. This increased forcing helps to move $\mathbf{x}_k$ in such a way that $\mathbf{H}\mathbf{x}_k$ is closer to $\mathbf{z}_k$. This observation is corroborated by the plots of the evolution of the four components of the control $\{\mathbf{u}_k\}$ given in Figs. 4a–d. It is evident from Fig. 4 that the magnitude of the initial values of the control increases as the parameter c is decreased.

A standard measure of the closeness of the jth component of the controlled and uncontrolled model solution with the jth component of the observations are given by

$$\mathrm{RMS}_{1j} = \left[\frac{1}{N}\sum_{k=0}^{N}(z_{k,j} - x_{k,j})^2\right]^{1/2} \tag{5.15}$$

and

$$\mathrm{RMS}_{2j} = \left[\frac{1}{N}\sum_{k=0}^{N}(z_{k,j} - \xi_{k,j})^2\right]^{1/2}. \tag{5.16}$$

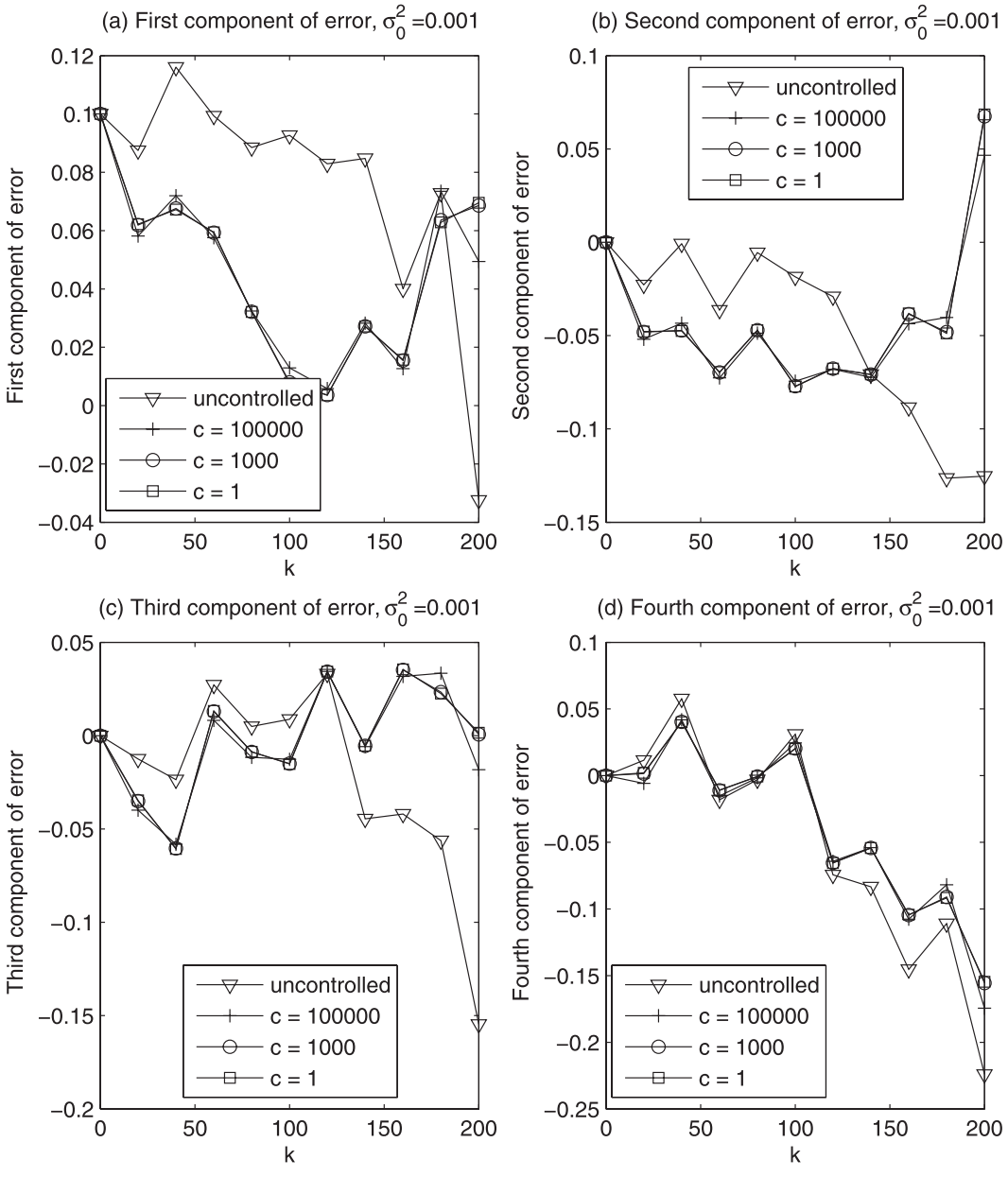

FIG. 5. Comparison of the four components of the uncontrolled error $e_0 = \boldsymbol{\xi}_k - \mathbf{z}_k$ and the controlled error $e_c = \mathbf{x}_k - \mathbf{z}_k$ for $p = 1$, $\mathbf{B} = (1\ 1\ 1\ 1)^{\mathrm{T}}$, and $c = \{100\ 000, 1000, 1\}$.

Table 2 gives the values of these measures for various combinations of the values of σ_0^2 and c. It is clear from Fig. 3 and Table 2 that for a given σ_0^2, RMS_1 decreases as c decreases.

2) EXPERIMENT 2

In this experiment we set $p = 1$ and $\mathbf{B} = (1\ 1\ 1\ 1)^{\mathrm{T}}$ and all the other parameters are the same as in experiment 1. A comparison of the plots of the observations with controlled and uncontrolled model solution is given in Fig. 5. Table 3 provides a comparison of the RMS errors for various choices of σ_0^2 and c. Recall that when $p = 1$, the same control is applied to every component of the state vector as opposed to when $p = 4$ where different elements of the control vector impact the evolution of the different components of the state vector. Thus in experiment 1 ($p = 4$) the components of the control vector are customized to each component of the state vector and hence the errors are less as borne by comparing the corresponding elements of Tables 2 and 3. Clearly, larger p is better.

TABLE 3. Root-mean-square errors of the controlled and uncontrolled model solution with observations [$p = 1, \mathbf{B} = (1\,1\,1\,1)^{\mathrm{T}}$].

σ_0^2	c	RMS_{e1}	RMS_{e2}	RMS_{e3}	RMS_{e4}
Uncontrolled		0.0850	0.0658	0.0551	0.0958
0.001	100 000	0.0539	0.0551	0.0286	0.0728
	1000	0.0546	0.0568	0.0275	0.0693
	1	0.0546	0.0570	0.0275	0.0691
0.005	100 000	0.0933	0.0869	0.0579	0.1010
	1000	0.0968	0.0821	0.0573	0.0962
	1	0.0973	0.0815	0.0573	0.0958
0.01	100 000	0.1000	0.1330	0.1410	0.1560
	1000	0.0846	0.1140	0.1430	0.1490
	1	0.0834	0.1110	0.1450	0.1490

6. Identification of model errors

One of the lofty goals of dynamic data assimilation is to find a correction for model error—errors due to the absence or inappropriate parameterization of physical processes germane to the phenomenon under investigation, and/or incorrect specification of the deterministic model's control vector (initial conditions, boundary conditions, and physical/empirical parameters). The theory developed in sections 2 and 3 and the illustrations in sections 4 and 5 demonstrate the inherent strength of Pontryagin's minimum principle as a means of finding this correction.

In an effort to further understand the sources of model error, we take the Pontryagin procedure one step further—we attempt to find a correction matrix $\mathbf{S} \in \mathbf{R}^{n\times n}$ such that the solution of the corrected but unforced model ($\mathbf{M} + \mathbf{S}$) matches the optimal trajectory from Pontryagin. That is,

$$\mathbf{x}_{k+1} = (\mathbf{M} + \mathbf{S})\mathbf{x}_k, \tag{6.1}$$

where $\{\mathbf{x}_k\}$ is the optimal trajectory of (5.3). Subtracting (5.3) from (6.1), we find

$$\mathbf{S}\mathbf{x}_k = \mathbf{y}_k \tag{6.2}$$

for all $1 \le k \le N$, where $\mathbf{y}_k = \mathbf{B}\mathbf{u}_k$. That is, given $\{\mathbf{x}_k\}$ and the optimal input time series $\{\mathbf{y}_k\}$, we seek to find a time invariant linear operator $\mathbf{S}$ that will map $\mathbf{x}_k$ to $\mathbf{y}_k$ for all $1 \le k \le N$.

This inverse problem can be recast as an unconstrained minimization of the quadratic functional $\mathbf{Q}$: $\mathbf{R}^{n\times n} \to \mathbf{R}$ defined by

$$\mathbf{Q}(\mathbf{S}) = \sum_{k=1}^{N} \mathbf{Q}_k(\mathbf{S}) \tag{6.3}$$

with respect to $\mathbf{S} \in \mathbf{R}^{n\times n}$, where

$$\mathbf{Q}_k(\mathbf{S}) = \frac{1}{2}(\mathbf{S}\mathbf{x}_k - \mathbf{y}_k)^{\mathrm{T}}(\mathbf{S}\mathbf{x}_k - \mathbf{y}_k)$$
$$= \frac{1}{2}[\boldsymbol{\alpha}(\mathbf{S}, \mathbf{x}_k) - 2\boldsymbol{\beta}(\mathbf{S}, \mathbf{x}_k, \mathbf{y}_k) + \boldsymbol{\gamma}(\mathbf{y}_k)] \tag{6.4}$$

and

$$\boldsymbol{\alpha}(\mathbf{S}, \mathbf{x}) = \mathbf{x}^{\mathrm{T}}(\mathbf{S}^{\mathrm{T}}\mathbf{S})\mathbf{x}, \tag{6.5}$$

$$\boldsymbol{\beta}(\mathbf{S}, \mathbf{x}, \mathbf{y}) = \mathbf{y}^{\mathrm{T}}\mathbf{S}\mathbf{x}, \tag{6.6}$$

$$\boldsymbol{\gamma}(\mathbf{y}) = \mathbf{y}^{\mathrm{T}}\mathbf{y}. \tag{6.7}$$

From appendix C it readily follows that the optimal $\mathbf{S}$ is given by

$$\mathbf{S} = \left[\sum_{k=1}^{N} \mathbf{y}_k\mathbf{x}_k^{\mathrm{T}}\right]\left[\sum_{k=1}^{N} \mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}\right]^{+}, \tag{6.8}$$

where $\mathbf{A}^{+}$ denotes the generalized inverse of $\mathbf{A}$.

Those familiar with optimal interpolation method (Gandin 1965) will readily recognize that the first term on the right-hand side of (6.8) is akin to the cross covariance between $\mathbf{x}_k$ and $\mathbf{y}_k$ and the second term is akin to the inverse of the covariance of $\mathbf{x}_k$ with itself. We now illustrate this idea in the following example.

Example 6.1

Using the optimal control sequence $\mathbf{y}_k = \mathbf{B}\mathbf{u}_k$ and its associated optimal trajectory $\mathbf{x}_k$ found in example 5.1 (with $n = 4$), the value of $\mathbf{S}$ computed from (6.8) is given by

$$\mathbf{S} = \begin{bmatrix} -0.0045 & -0.0176 & -0.0186 & -0.0074 \\ -0.0004 & -0.0006 & 0.0063 & 0.0084 \\ 0.0009 & 0.0043 & 0.0009 & -0.0081 \\ -0.0009 & -0.0011 & -0.0001 & -0.0173 \end{bmatrix}. \tag{6.9}$$

The trajectory of the corrected but uncontrolled model is given by

$$\boldsymbol{\zeta}_{k+1} = (\mathbf{M} + \mathbf{S})\boldsymbol{\zeta}_k, \quad \boldsymbol{\zeta}_0 = \mathbf{x}_0. \tag{6.10}$$

A comparison of $\boldsymbol{\zeta}_k - \mathbf{z}_k$, the error between corrected but uncontrolled model in (6.10), and $\boldsymbol{\xi}_k - \mathbf{z}_k$, the error between the uncorrected and uncontrolled model in (5.11), is given in Fig. 6. It is evident from Fig. 6 that the trajectory of the corrected but uncontrolled model fits the observations better.

We conclude this section with the following remarks:

1) Define a vector $\mathbf{s}(\mathbf{x}) = (\sin\mathbf{x}, \sin2\mathbf{x}, \sin3\mathbf{x}, \sin4\mathbf{x})^{\mathrm{T}}$ and define

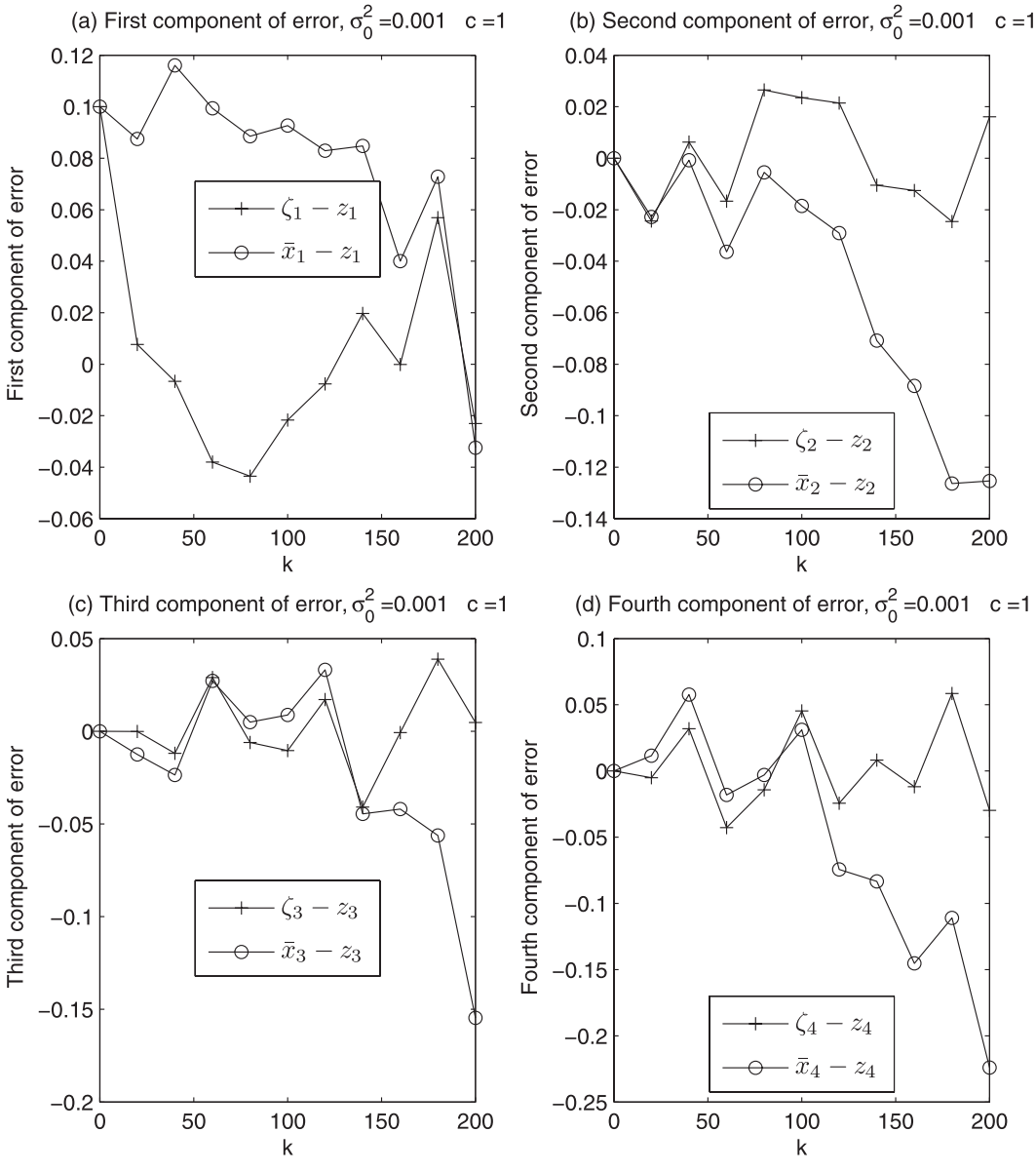

FIG. 6. Comparison of the four components of the error $\boldsymbol{\zeta}_k - \mathbf{z}_k$ between the corrected but uncontrolled model state $\{\boldsymbol{\zeta}_k\}$ in (6.10) with the observation $\{\mathbf{z}_k\}$ and the error $\bar{\mathbf{x}}_k - \mathbf{z}_k$ between the original uncorrected and uncontrolled model state $\{\bar{\mathbf{x}}_k\}$ in (2.1) with the observation $\{\mathbf{z}_k\}$.

$$\mathbf{q}_1(\mathbf{x}, k) = \boldsymbol{\xi}_k^{\mathrm{T}}\mathbf{s}(\mathbf{x})$$
$$\mathbf{q}_2(\mathbf{x}, k) = \boldsymbol{\zeta}_k^{\mathrm{T}}\mathbf{s}(\mathbf{x}), \tag{6.11}$$

where $\boldsymbol{\xi}_k$ is the (uncontrolled) model trajectory obtained from (5.1) using matrix $\mathbf{M}$ and $\boldsymbol{\zeta}_k$ is the (uncontrolled) model trajectory obtained from (6.10) with matrix $(\mathbf{M} + \mathbf{S})$. Clearly $\mathbf{q}_1(\mathbf{x}, k)$ and $\mathbf{q}_2(\mathbf{x}, k)$ are approximations to the exact solution $\mathbf{q}(\mathbf{x}, t)$ in (4.3) at $t = k\Delta t$. It can be verified that

$$|\mathbf{q}(\mathbf{x}, k) - \mathbf{q}_2(\mathbf{x}, k)| \le |\mathbf{q}(\mathbf{x}, k) - \mathbf{q}_1(\mathbf{x}, k)|, \tag{6.12}$$

where $\mathbf{q}(\mathbf{x}, k) = \mathbf{q}(\mathbf{x}, t)$ at $t = k\Delta t$. That is, the one-step model error correlation matrix $\mathbf{S}$ forces the model solution closer to the true solution.

2) Only for simplicity in exposition did we pose the inverse problem in (6.3) as an unconstrained problem. In fact, one could readily accommodate structural constraint on $\mathbf{S}$—such as requiring it to be a diagonal, tridiagonal, or lower-triangular matrix, etc. Further, we could also readily impose inequality constraints on a selected subset of elements of $\mathbf{S}$.

3) Again, only for simplicity did we obtain a single matrix $\mathbf{S}$ that covers the entire assimilation window and mapping $\mathbf{x}_k$ to $\mathbf{y}_k$ for all $1 \le k \le N$. In principle, we could divide the assimilation window into L subintervals. Then we could estimate the correction matrix $\mathbf{S}_q$ using only the $(\mathbf{x}_k, \mathbf{y}_k)$ pairs that reside in the qth subinterval. In this latter case, we will have a time varying one-step transition correction matrix $\mathbf{S}_q$ for each subinterval, $1 \le q \le L$.

7. Conclusions

The essence of the PMP-based approach to dynamic data assimilation is computation of optimal control sequence $\mathbf{u}_k \in \mathbf{R}^p$ where the parameter $1 \le p \le n$ denotes the ''richness'' of the control. It follows from experiments 1 and 2 that a larger value of p is better. And when this sequence is applied to the deterministic model, it forces the model to track the observations as closely as desired where the closeness is controlled by judicious choices of the relative weights of the two energy terms in the cost functional. More specifically, for a given observational noise covariance matrix $\mathbf{R}$, a simple choice of $\mathbf{C} = c\mathbf{I}_p$ with smaller value of the constant c provides a better fit between the model and the data. The computation of this optimal control sequence rests on the solution to a nonlinear TPBVP. While the solution to this latter class of problems can be a daunting task, especially for the large-scale problems of interest in the geosciences, several well-tested numerical methods for finding the solution are known and are available as components of several program libraries.

We have demonstrated the power of this approach by applying it to a nontrivial linear advection problem. For this linear problem, the TPBVP reduces to two initial value problems. In addition we have developed a flexible framework to consolidate the information from the optimal control sequence into a single correction matrix $\mathbf{S}$, which, when added to the given model matrix $\mathbf{M}$, will indeed generate a solution that will closely match the optimal trajectory computed using the PMP.

It should be interesting and valuable to compare the model error corrections obtained using the PMP with those obtained from using the model in a weak constraint formulation.

Acknowledgments. We are very grateful to Qin Xu and an anonymous reviewer for their comments and suggestions that helped to improve the organizations of the paper. S. Lakshmivarahan's efforts are supported in part by two grants: NSF EPSCOR Track 2 Grant 105-155900 and NSF Grant 105-15400.

APPENDIX A

On the Correctness of the Affine Relation between the Costate Variable $\boldsymbol{\lambda}_k$ and the State Variable $\mathbf{x}_k$ Given in (3.8)

Since $\boldsymbol{\lambda}_N = 0$, from the second equation in (3.7) we get

$$\boldsymbol{\lambda}_{N-1} = \mathbf{F}\mathbf{x}_{N-1} - \mathbf{W}\mathbf{z}_{N-1}, \tag{A.1}$$

which clearly shows that $\boldsymbol{\lambda}_{N-1}$ is an affine function of $\mathbf{x}_{N-1}$. Substituting (A.1) back into the second equation in (3.7), we obtain

$$\boldsymbol{\lambda}_{N-2} = \mathbf{F}\mathbf{x}_{N-2} + \mathbf{M}^{\mathrm{T}}(\mathbf{F}\mathbf{x}_{N-1} - \mathbf{W}\mathbf{z}_{N-1}) - \mathbf{W}\mathbf{z}_{N-2}. \tag{A.2}$$

But from the first equation in (3.7), it follows that

$$\mathbf{x}_{N-1} = \mathbf{M}\mathbf{x}_{N-2} - \mathbf{E}\boldsymbol{\lambda}_{N-1}. \tag{A.3}$$

Using (A.1) in (A.3) and simplifying we get

$$\mathbf{x}_{N-1} = (\mathbf{I} + \mathbf{E}\mathbf{F})^{-1}(\mathbf{M}\mathbf{x}_{N-2} + \mathbf{E}\mathbf{W}\mathbf{z}_{N-1}). \tag{A.4}$$

Now substituting (A.4) into (A.2), it follows that

$$\begin{aligned}\boldsymbol{\lambda}_{N-2} = {} & [\mathbf{F} + \mathbf{M}^{\mathrm{T}}\mathbf{F}(\mathbf{I} + \mathbf{E}\mathbf{F})^{-1}\mathbf{M}]\mathbf{x}_{N-2} \\ & + \mathbf{M}^{\mathrm{T}}(\mathbf{I} + \mathbf{E}\mathbf{F})^{-1}\mathbf{E}\mathbf{W}\mathbf{z}_{N-1} - \mathbf{M}^{\mathrm{T}}\mathbf{W}\mathbf{z}_{N-1} - \mathbf{W}\mathbf{z}_{N-2},\end{aligned} \tag{A.5}$$

which is clearly affine in $\mathbf{x}_{N-2}$.

Continuing inductively it can be easily verified that $\boldsymbol{\lambda}_k$ is an affine function of $\mathbf{x}_k$ as posited in (3.8).

APPENDIX B

Solution of the LOM(n) in (4.6)

In this appendix we analyze the eigenstructure of the matrix $\mathbf{A}$ in (4.8) leading to its Jordan canonical form, which, in turn, leads to the closed form solution of the LOM in (4.8).

a. Eigenstructure of the matrix $\mathbf{A}$

Since the structure of the matrix $\mathbf{A}$ in (4.8) is closely related to the tridiagonal matrix, we start this discussion by stating a well-known result relating to the recursive computation of the determinant of the tridiagonal matrix (Lakshmivarahan and Dhall 1990, 416–418).

Let $\mathbf{B}_k \in \mathbf{R}^{k\times k}$ be a general tridiagonal matrix of the form

TABLE B1. Determinant, characteristic polynomial, and eigenvalues of the matrix $\mathbf{A}$ for $2 \le n \le 10$.

n	Determinant $\lvert\mathbf{A}\rvert = n!/2^n$ (n even)	Characteristic polynomial of $\mathbf{A}$	Eigenvalues of $\mathbf{A}$
2	1/2	$\boldsymbol{\lambda}^2 + 1/2$	$\pm i(1/\sqrt{2})$
3	0	$\lambda(\lambda^2 + 2)$	$0, \pm\sqrt{2}$
4	3/2	$\boldsymbol{\lambda}^4 + 5\boldsymbol{\lambda}^2 + 3/2$	$\pm i(2.1632), \pm i(0.5662)$
5	0	$\boldsymbol{\lambda}(\boldsymbol{\lambda}^4 + 10\boldsymbol{\lambda}^2 + 23/2)$	$0, \pm i(2.9452), \pm i(1.1514)$
6	45/4	$\boldsymbol{\lambda}^6 + (35/2)\boldsymbol{\lambda}^4 + 49\boldsymbol{\lambda}^2 + 45/4$	$\pm i(3.7517), \pm i(1.7812), \pm i(0.5019)$
7	0	$\boldsymbol{\lambda}(\boldsymbol{\lambda}^6 + 28\boldsymbol{\lambda}^4 + 154\boldsymbol{\lambda}^2 + 132)$	$0, \pm i(4.5771), \pm i(2.4495), \pm i(1.0297)$
8	315/2	$\boldsymbol{\lambda}^8 + 42\boldsymbol{\lambda}^6 + 399\boldsymbol{\lambda}^4 + 818\boldsymbol{\lambda}^2 + 315/2$	$\pm i(5.4174), \pm i(3.1486), \pm i(1.5937), \pm i(0.4631)$
9	0	$\boldsymbol{\lambda}(\boldsymbol{\lambda}^8 + 60\boldsymbol{\lambda}^6 + 903\boldsymbol{\lambda}^4 + 3590\boldsymbol{\lambda}^2) + 5067/2$	$0, \pm i(6.2698), \pm i(3.8730), \pm i(2.1906), \pm i(0.9460)$
10	14 175/4	$\boldsymbol{\lambda}^{10} + 165\boldsymbol{\lambda}^8 + 1848\boldsymbol{\lambda}^6 + (25\,235/2)\boldsymbol{\lambda}^4 + (41\,877/2)\boldsymbol{\lambda}^2 + 14\,175/4$	$\pm i(7.1323), \pm i(4.6165), \pm i(2.8239), \pm i(1.5860), \pm i(0.4363)$

$$\mathbf{B}_k = \begin{bmatrix} b_1 & c_1 & 0 & . & . & . & . & 0 & 0 & 0 \\ a_2 & b_2 & c_2 & 0 & . & . & . & 0 & 0 & 0 \\ 0 & a_3 & b_3 & c_3 & . & . & . & 0 & 0 & 0 \\ . & . & . & . & & & & . & . & . \\ . & . & . & . & & & & . & . & . \\ . & . & . & . & & & & . & . & . \\ . & . & . & . & & & & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & . & . & . & a_{k-1} & b_{k-1} & c_{k-1} \\ 0 & 0 & 0 & 0 & . & . & . & 0 & a_k & b_k \end{bmatrix} \tag{B.1}$$

for $1 \le k \le n$. Let $\mathbf{D}_i$ denote the determinant of the principal submatrix consisting of the first i rows and i columns of $\mathbf{B}_k$. Then the determinant $\mathbf{D}_k$ of the matrix $\mathbf{B}_k$ in (B.1) is obtained by applying the Laplace expansion to the kth (last) row of $\mathbf{B}_k$ and is given by the second-order linear recurrence

$$\mathbf{D}_k = b_k\mathbf{D}_{k-1} - a_kc_{k-1}\mathbf{D}_{k-2}, \tag{B.2}$$

where $\mathbf{D}_0 = 1$ and $\mathbf{D}_1 = b_1$.

Setting $b_i \equiv 0$ for all $1 \le i \le n$, $c_i = (1/2)(i + 1)$ for $1 \le i \le (n - 1)$, and $a_i = -(1/2)(i - 1)$ for $2 \le i \le n$ in (B.1), it can be verified that $\mathbf{B}_n$ reduces to $\mathbf{A}$ in (4.10). Substituting these values in (B.2), the latter becomes

$$\mathbf{D}_k = 0 \cdot \mathbf{D}_{k-1} + \frac{k(k-1)}{4}\mathbf{D}_{k-2}, \tag{B.3}$$

with $\mathbf{D}_0 = 1$ and $\mathbf{D}_1 = 0$. Iterating (B.3), it can be verified that

$$\mathbf{D}_k = \begin{cases} \dfrac{k!}{2^k} & \text{if} \quad k \text{ is even} \\ 0 & \text{if} \quad k \text{ is odd} \end{cases}. \tag{B.4}$$

Thus, $\mathbf{A}$ in (4.8) is singular when n is odd. Henceforth we only consider the case when n is even. Refer to Table B1 for values of the determinant of $\mathbf{A}$ for $2 \le n \le 10$.

The characteristic polynomial of $\mathbf{A}$ in (B.1) is found by setting $b_i = -\boldsymbol{\lambda}$, $c_i = (1/2)(i + 1)$, and $a_i = -(1/2)(i - 1)$. In this case, the determinant $\mathbf{D}_n$ of $\mathbf{B}_n$ in (B.1) represents the characteristic polynomial of $\mathbf{A}$ in (4.8). Making the above substitutions in (B.2), the latter becomes

$$\mathbf{D}_k = -\boldsymbol{\lambda} \cdot \mathbf{D}_{k-1} + \frac{k(k-1)}{4}\mathbf{D}_{k-2}, \tag{B.5}$$

with $\mathbf{D}_0 = 1$ and $\mathbf{D}_1 = -\boldsymbol{\lambda}$. Iterating (B.5) leads to the sequence of characteristic polynomials of $\mathbf{A}$ for various values of n. Table B1 also gives the characteristic polynomial and the eigenvalues of $\mathbf{A}$ for $2 \le n \le 10$. From this table it is clear that the absolute value of the largest eigenvalue increases and that of the smallest (nonzero) eigenvalue decreases with n. It turns out that the larger eigenvalues correspond to the high-frequency components and the smaller eigenvalues correspond to the low-frequency components that make up the solution $\mathbf{q}(t)$ of (4.8).

b. Jordan canonical form for $\mathbf{A}$

Let $\boldsymbol{\Lambda} \in \mathbf{R}^{n\times n}$ denote the matrix eigenvalues of $\mathbf{A}$ and let $\mathbf{V} \in \mathbf{R}^{n\times n}$ denote a nonsingular matrix of the corresponding eigenvectors; that is,

$$\mathbf{AV} = \mathbf{V}\boldsymbol{\Lambda}. \tag{B.6}$$

Then

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1} \quad \text{and} \quad \mathbf{V}^{-1}\mathbf{AV} = \boldsymbol{\Lambda}, \tag{B.7}$$

and $\boldsymbol{\Lambda}$ takes a special block diagonal form

$$\boldsymbol{\Lambda} = \begin{bmatrix} \mathbf{L}_1 & 0 & . & . & . & 0 \\ 0 & \mathbf{L}_2 & 0 & . & . & 0 \\ . & 0 & \mathbf{L}_3 & 0 & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & \mathbf{L}_{n/2} \end{bmatrix} \tag{B.8}$$

and

$$\mathbf{L}_i = \begin{bmatrix} 0 & \boldsymbol{\lambda}_i \\ -\boldsymbol{\lambda}_i & 0 \end{bmatrix} \tag{B.9}$$

for each complex conjugate pair $\pm i\boldsymbol{\lambda}_i$ of eigenvalues of $\mathbf{A}$ for $1 \le i \le (n/2)$. The matrix $\boldsymbol{\Lambda}$ in (B.8) is known as the Jordan canonical form of $\mathbf{A}$ (Hirsch and Smale 1974).

SOLUTION OF (4.8):

The general form of the solution $\mathbf{q}(t)$ of (4.8) is given by

$$\mathbf{q}(t) = e^{\mathbf{A}t}\mathbf{q}(0). \tag{B.10}$$

Using (B.7) in (B.10), it can be shown that

$$\mathbf{q}(t) = e^{(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1})t}\mathbf{q}(0) = \mathbf{V}e^{\boldsymbol{\Lambda}t}\mathbf{V}^{-1}\mathbf{q}(0) \tag{B.11}$$

or

$$\overline{\mathbf{q}}(t) = e^{\boldsymbol{\Lambda}t}\overline{\mathbf{q}}(0), \tag{B.12}$$

where $\mathbf{q}(t)$ and $\overline{\mathbf{q}}(t)$ are related by the linear transformation

$$\overline{\mathbf{q}}(t) = \mathbf{V}^{-1}\mathbf{q}(t). \tag{B.13}$$

By exploiting the structure of $\boldsymbol{\Lambda}$, it can be verified that

$$e^{\boldsymbol{\Lambda}t} = \begin{bmatrix} e^{\mathbf{L}_1 t} & & & \\ & e^{\mathbf{L}_2 t} & & \\ & & \cdots & \\ & & & e^{\mathbf{L}_{n/2} t} \end{bmatrix}, \tag{B.14}$$

where

$$e^{\mathbf{L}_i t} = \begin{bmatrix} c_i & s_i \\ -s_i & c_i \end{bmatrix} \tag{B.15}$$

and

$$c_i = \cos(\boldsymbol{\lambda}_i t) \quad \text{and} \quad s_i = \sin(\boldsymbol{\lambda}_i t). \tag{B.16}$$

Substituting (B.14)–(B.16) into (B.12), we obtain $\overline{\mathbf{q}}(t)$. Clearly, $\mathbf{q}(t) = \mathbf{V}\overline{\mathbf{q}}(t)$ is the solution of (4.8).

We conclude this appendix with the following.

Example (B.1). Consider the case with $n = 4$ and

$$\frac{d\mathbf{q}(t)}{dt} = \mathbf{A}\mathbf{q}(t), \tag{B.17}$$

with $\mathbf{A}$ given by (4.8). From Table B1, the eigenvalues of $\mathbf{A}$ (listed in the increasing order of their absolute values computed using MATLAB) are given by

$$\pm i\boldsymbol{\lambda}_1 = \pm i(0.5662) \quad \text{and} \quad \pm i\boldsymbol{\lambda}_2 = \pm i(2.1662). \tag{B.18}$$

From (B.14)–(B.16), we obtain

$$\mathbf{L}_1 = \begin{bmatrix} 0 & 0.5662 \\ -0.5662 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{L}_2 = \begin{bmatrix} 0 & 2.1662 \\ -2.1662 & 0 \end{bmatrix} \tag{B.19}$$

and

$$e^{\boldsymbol{\Lambda}t} = \begin{bmatrix} e^{\mathbf{L}_1 t} & 0 \\ 0 & e^{\mathbf{L}_2 t} \end{bmatrix}, \tag{B.20}$$

where

$$e^{\mathbf{L}_1 t} = \begin{bmatrix} c_1 & s_1 \\ -s_1 & c_1 \end{bmatrix} \quad \text{and} \quad e^{\mathbf{L}_2 t} = \begin{bmatrix} c_2 & s_2 \\ -s_2 & c_2 \end{bmatrix} \tag{B.21}$$

and

$$c_1 = \cos(0.5662t), \quad c_2 = \cos(2.1662t),$$
$$s_1 = \sin(0.5662t), \quad s_2 = \sin(2.1662t).$$

Hence, $\overline{\mathbf{q}}(t) \in \mathbf{R}^4$ is given by

$$\begin{aligned} \overline{\mathbf{q}}_1(t) &= c_1\overline{\mathbf{q}}_1(0) + s_1\overline{\mathbf{q}}_2(0) \\ \overline{\mathbf{q}}_2(t) &= -s_1\overline{\mathbf{q}}_1(0) + c_1\overline{\mathbf{q}}_2(0) \\ \overline{\mathbf{q}}_3(t) &= c_2\overline{\mathbf{q}}_3(0) + s_2\overline{\mathbf{q}}_4(0) \\ \overline{\mathbf{q}}_4(t) &= -s_2\overline{\mathbf{q}}_3(0) + c_2\overline{\mathbf{q}}_4(0). \end{aligned} \tag{B.22}$$

It can be verified that the matrix of eigenvector $\mathbf{V}$ corresponding to $\boldsymbol{\Lambda}$ above is given by

$$\mathbf{V} = \begin{bmatrix} -0.8340 & 0 & -0.2413 & 0 \\ 0 & 0.4726 & 0 & 0.5220 \\ -0.0999 & 0 & 0.6723 & 0 \\ 0 & 0.2646 & 0 & -0.4662 \end{bmatrix}. \tag{B.23}$$

Hence, the solution of (B.17) is given by

$$\begin{aligned} \mathbf{q}_1(t) &= -0.8340 \cdot \overline{\mathbf{q}}_1(t) - 0.2413 \cdot \overline{\mathbf{q}}_3(t) \\ \mathbf{q}_2(t) &= 0.4726 \cdot \overline{\mathbf{q}}_2(t) + 0.5220 \cdot \overline{\mathbf{q}}_4(t) \\ \mathbf{q}_3(t) &= -0.0999 \cdot \overline{\mathbf{q}}_1(t) + 0.6723 \cdot \overline{\mathbf{q}}_3(t) \\ \mathbf{q}_4(t) &= 0.2646 \cdot \overline{\mathbf{q}}_2(t) - 0.4662 \cdot \overline{\mathbf{q}}_4(t). \end{aligned} \tag{B.24}$$

Clearly, the general solution $\mathbf{q}_i(t)$ for each i is a linear combination of the harmonic terms $\cos(\boldsymbol{\lambda}_k t)$ and $\sin(\boldsymbol{\lambda}_k t)$, $1 \le k \le n/2$, where the coefficients of the linear combination are given by the elements of the ith row of the matrix $\mathbf{V}$ of eigenvectors of $\mathbf{A}$.

APPENDIX C

Gradient of Q(S) in (6.3)

Let $\mathbf{Q}$: $\mathbf{R}^{n\times n} \rightarrow \mathbf{R}$ be a functional defined over a set of $n \times n$ matrices. Then, by definition, the gradient $\nabla_{\mathbf{S}}\mathbf{Q}(\mathbf{S})$ is a matrix given by

$$\nabla_{\mathbf{S}}\mathbf{Q}(\mathbf{S}) = \left[\frac{\partial \mathbf{Q}(\mathbf{S})}{\partial \mathbf{S}_{ij}}\right]. \tag{C.1}$$

For the gradient of $\alpha(\mathbf{S}, \mathbf{x})$ in (6.5), let

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{1^*} \\ \mathbf{S}_{2^*} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{S}_{n^*} \end{bmatrix} \tag{C.2}$$

be a row partition of $\mathbf{S}$. Then, the Grammian $\mathbf{S}^{\mathrm{T}}\mathbf{S}$ can be expressed as

$$\mathbf{S}^{\mathrm{T}}\mathbf{S} = \sum_{i=1}^{n} \mathbf{S}_{i^*}^{\mathrm{T}}\mathbf{S}_{i^*}. \tag{C.3}$$

Consequently,

$$\alpha(\mathbf{S}, \mathbf{x}) = \mathbf{x}^{\mathrm{T}}(\mathbf{S}^{\mathrm{T}}\mathbf{S})\mathbf{x} = \sum_{i=1}^{n} \mathbf{x}^{\mathrm{T}}(\mathbf{S}_{i^*}^{\mathrm{T}}\mathbf{S}_{i^*})\mathbf{x} = \sum_{i=1}^{n} (\mathbf{S}_{i^*}\mathbf{x})^2. \tag{C.4}$$

Hence the gradient of $\alpha(\mathbf{S}, \mathbf{x})$ with respect to the column vector $\mathbf{S}_{i^*}^{\mathrm{T}}$ is given by

$$\nabla_{\mathbf{S}_{i^*}^{\mathrm{T}}}\alpha(\mathbf{S}, \mathbf{x}) = 2(\mathbf{x}\mathbf{x}^{\mathrm{T}})\mathbf{S}_{i^*}^{\mathrm{T}}. \tag{C.5}$$

Taking transpose of both sides,

$$\nabla_{\mathbf{S}_{i^*}}\alpha(\mathbf{S}, \mathbf{x}) = 2\mathbf{S}_{i^*}(\mathbf{x}\mathbf{x}^{\mathrm{T}}). \tag{C.6}$$

By stacking these rows of derivatives, we get

$$\nabla_{\mathbf{S}}\alpha(\mathbf{S}, \mathbf{x}) = 2\mathbf{S}(\mathbf{x}\mathbf{x}^{\mathrm{T}}). \tag{C.7}$$

a. Gradient of $\boldsymbol{\beta}(\mathbf{S}, \mathbf{x}, \mathbf{y})$ in (6.6)

From (6.6),

$$\boldsymbol{\beta}(\mathbf{S}, \mathbf{x}, \mathbf{y}) = \mathbf{y}^{\mathrm{T}}\mathbf{S}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{y}_i\mathbf{S}_{ij}\mathbf{x}_j. \tag{C.8}$$

Hence,

$$\nabla_{\mathbf{S}}\boldsymbol{\beta}(\mathbf{x}, \mathbf{y}) = \mathbf{y}\mathbf{x}^{\mathrm{T}}. \tag{C.9}$$

b. Gradient of $\mathbf{Q}(\mathbf{S})$ in (6.3)

Combining (C.7) and (C.9) with (6.4)–(6.7), it is immediate that

$$\nabla_{\mathbf{S}}\mathbf{Q}_k(\mathbf{S}) = \mathbf{S}(\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}) - \mathbf{y}_k\mathbf{x}_k^{\mathrm{T}} \tag{C.10}$$

and

$$\nabla_{\mathbf{S}}\mathbf{Q}(\mathbf{S}) = \mathbf{S}\sum_{k=1}^{N}(\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}) - \sum_{k=1}^{N}(\mathbf{y}_k\mathbf{x}_k^{\mathrm{T}}). \tag{C.11}$$

Hence, the minimizer of $\mathbf{Q}(\mathbf{S})$ in (6.3) is given by

$$\mathbf{S} = \left[\sum_{k=1}^{N} \mathbf{y}_k\mathbf{x}_k^{\mathrm{T}}\right]\left[\sum_{k=1}^{N} \mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}\right]^{+}, \tag{C.12}$$

where $\mathbf{A}^{+}$ is the generalized inverse of $\mathbf{A}$.

REFERENCES

Abramov, R. V., G. Kovacic, and A. J. Majda, 2003: Hamiltonian structure and statistically relevant conserved quantities for the truncated Burger-Hopf equation. *Commun. Pure Appl. Math.,* **56,** 1–46.

Anthes, R. A., 1974: Data assimilation and initialization of hurricane prediction model. *J. Atmos. Sci.,* **31,** 702–719.

Athans, M., and P. L. Falb, 1966*: Optimal Control.* McGraw-Hill, 879 pp.

Bennett, A., 1992: *Inverse Methods in Physical Oceanography.* Cambridge University Press, 346 pp.

——, and M. A. Thorburn, 1992: The generalized inverse of a nonlinear quasigeostrophic ocean circulation model. *J. Phys. Oceanogr.,* **22,** 213–230.

Bennett, S., 1996: A brief history of automatic control. *IEEE Control Syst.,* **16,** 17–25.

Bergman, K. H., 1979: Multivariate analysis of temperatures and winds using optimum interpolation. *Mon. Wea. Rev.,* **107,** 1423–1444.

Bergthórsson, P., and B. Döös, 1955: Numerical weather map analysis. *Tellus,* **7,** 329–340.

Boltyanskii, V. G., 1971: *Mathematical Methods of Optimal Control.* Holt, Rinehart and Winston, 272 pp.

——, 1978: *Optimal Control of Discrete Systems.* John Wiley and Sons, 392 pp.

Bryson, A. E., 1996: Optimal control-1950 to 1985. *IEEE Control Syst.,* **16,** 26–33.

——, 1999: *Dynamic Optimization.* Addison-Wesley, 434 pp.
Canon, M. D., C. D. Cullum Jr., and E. Polak, 1970: *Theory of Optimal Control and Mathematical Programming.* McGraw Hill, 285 pp.
Carrier, G. F., and C. E. Pearson, 1976: *Partial Differential Equations: Theory and Techniques.* Academic Press, 320 pp.
Catlin, D. E., 1989: *Estimation, Control and the Discrete Kalman Filter.* Springer-Verlag, 274 pp.
Dee, D. P., and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.,* **124,** 269–295.
Derber, J., 1989: A variational continuous assimilation technique. *Mon. Wea. Rev.,* **117,** 2437–2446.
Eliassen, A., 1954: Provisional report on the calculation of spatial covariance and autocorrelation of pressure field. Institute of Weather and Climate Research, Academy of Sciences Rep. 5, 12 pp. [Available from Norwegian Meteorological Institute, P.O. Box 43, Blindern, N-0313, Oslo, Norway.]
Friedland, B., 1969: Treatment of bias in recursive filtering. *IEEE Trans. Autom. Control.,* **14,** 359–367.
Gandin, L. S., 1965: *Objective Analysis of Meteorological Fields.* Israel Program for Scientific Translations, 242 pp.
Goldstein, H. H., 1950: *Classical Mechanics.* Addison-Wesley, 399 pp.
——, 1980: *A History of the Calculus of Variations from the 17th through the 19th Century.* Springer-Verlag, 410 pp.
Griffith, A. K., and N. K. Nichols, 2001: Adjoint methods in data assimilation for estimating model error. *Flow, Turbul. Combust.,* **65,** 469–488.
Hirsch, M. W., and S. Smale, 1974: *Differential Equations, Dynamical Systems, and Linear Algebra.* Academic Press, 358 pp.
Kalman, R. E., 1963: The theory of optimal control and calculus of variations. *Mathematical Optimization Techniques,* R. Bellman, Ed., University of California Press, 309–329.
Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation, and Predictability.* Cambridge University Press, 341 pp.
Keller, H. B., 1976: *Numerical Solution of Two Point Boundary Value Problems.* Regional Conference Series in Applied Mathematics, Vol. 24, SIAM Publications, 69 pp.
Kuhn, H. W., and A. W. Tucker, 1951: Nonlinear programming. *Proc. Second Berkeley Symp. on Mathematical Statistics and Probability,* Berkeley, CA, University of California, Berkeley, 481–492.
Lakshmivarahan, S., and S. K. Dhall, 1990: *Analysis and Design of Parallel Algorithm: Arithmetic and Matrix Problems.* McGraw Hill, 657 pp.
——, and J. M. Lewis, 2013: Nudging: A critical overview. *Data Assmilation for Atmospheric, Oceanic and Hydrologic Applications,* Vol. 2, S. K. Park and L. Liang, Eds., Springer-Verlag, in press.
Lewis, F. L., 1986: *Optimal Control.* John Wiley and Sons, 362 pp.
Lewis, J. M., 1972: An operational upper air analysis using the variational methods. *Tellus,* **24,** 514–530.
——, and S. Lakshmivarahan, 2008: Sasaki's pivotal contribution: Calculus of variation applied to weather map analysis. *Mon. Wea. Rev.,* **136,** 3553–3567.
——, ——, and S. K. Dhall, 2006: *Dynamic Data Assimilation: A Least Squares Approach.* Cambridge University Press, 654 pp.
Lynch, P., 2006: *The Emergence of Numerical Weather Prediction: Richardson's Dream.* Cambridge University Press, 279 pp.
Majda, A. J., and I. Timofeyev, 2000: Remarkable statistical behavior for truncated Burgers–Hopf dynamics. *Proc. Natl. Acad. Sci. USA,* **97,** 12 413–12 417.
——, and ——, 2002: Statistical mechanics for truncations of Burger-Hopf equation: A model for intrinsic stochastic behavior with scaling. *Milan J. Math.,* **70,** 39–96.
Menard, R., and R. Daley, 1996: The application of Kalman smoother theory to estimation of 4DVAR error statistics. *Tellus,* **48A,** 221–237.
Naidu, D. S., 2003: *Optimal Control Systems.* CRC Press, 433 pp.
Platzman, G. W., 1964: An exact integral of complete spectral equations for unsteady one-dimensional flow. *Tellus,* **16,** 422–431.
Polak, E., 1997: *Optimization.* Springer, 779 pp.
Pontryagin, L. S., V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mischenko, 1962*: The Mathematical Theory of Optimal Control Processes.* John Wiley, 360 pp.
Roberts, S. M., and J. S. Shipman, 1972: *Two-Point Boundary Value Problems: Shooting Method.* Elsevier, 289 pp.
Rouch, H. E., F. Tung, and C. T. Striebel, 1965: Maximum likelihood estimates of linear dynamic systems. *J. Amer. Inst. Aeronaut. Astronaut.,* **3,** 1445–1450.
Sasaki, Y., 1958: An objective analysis based on the variational method. *J. Meteor. Soc. Japan,* **36,** 77–88.
——, 1970a: Some basic formulations in numerical variational analysis. *Mon. Wea. Rev.,* **98,** 875–883.
——, 1970b: Numerical variational analysis formulated under the constraints as determined by longwave equations and low-pass filter. *Mon. Wea. Rev.,* **98,** 884–898.
——, 1970c: Numerical variational analysis with weak constraint and application to surface analysis of severe storm gust. *Mon. Wea. Rev.,* **98,** 899–910.
Shen, J., T. Tang, and L. L. Wang, 2011: *Spectral Methods.* Springer-Verlag, 470 pp.
Wiener, N., 1948: *Cybernetics: Control and Communication in the Animal and Machine.* John Wiley, 194 pp.
Wiin-Nielsen, A., 1991: The birth of numerical weather prediction. *Tellus,* **43A,** 36–52.
Zupanski, D., 1997: A general weak constraint applicable to operational 4DVAR data assimilation system. *Mon. Wea. Rev.,* **125,** 2274–2292.